

Reforms on a knife edge

Frustration among French researchers at an apparent slowing down in promised reforms threatens to undermine an essential process. A less abrasive attitude on the government's part is essential.

France faces a unique opportunity to modernize its science base. Its scientific community — previously suspicious of reform — is not only clearly open to change but actively demanding it. There is a bristling optimism in the air, reminiscent of the high hopes that accompanied the election of President François Mitterrand in 1981. The then science minister, Jean-Pierre Chevènement, allowed this enthusiasm to express itself by organizing a whirlwind series of colloquia, the outcome of which was a white paper laying the foundations for a rejuvenation of French science during the 1980s.

Claude Allègre, the current minister of national education, research and technology, is well placed not only to match Chevènement's achievements but to surpass them. As a prominent Earth scientist with considerable administrative and political experience, his grasp of what science is about, and his vision of what it could deliver to society, probably surpasses that of any of his predecessors. And with a superministry responsible for virtually every aspect of science and technology, he has the means to put this vision into practice.

Allègre has already moved swiftly to redress one of the fundamental problems facing French science, its ageing population of researchers. The infusion of thousands of new posts — and a promise to put young researchers at the heart of all of his policies — is a wise step that should provide handsome benefits in the long term. His goal of reducing bureaucracy is also admirable. Continuing in the same vein, Allègre has also made it clear that he wants to put high-level research back into the universities, where it can nourish entire generations of students, consistent with his broader goal of injecting the

benefits of science into all corners of society.

In short, French science has at its helm a competent captain, who is charting a course that few would dispute is enlightened. Why, then, is discontent rumbling in the ranks of the scientific community, many of whom feel that the ship of reform, having raced down the boat-slip, is now idling in harbour instead of steaming out to sea (see page 8)?

Some of this impatience is unfair. French science, with its massive research organizations and technological programmes, is not an agile speedboat but a supertanker that needs time to turn. Changing direction, however, and more importantly steaming ahead, also requires having the crew on one's side. Past science ministers have often been paralysed by the fact that attempts to shake French science out of its leisurely cruise usually end in mutiny. Allègre has made it clear that he is having none of this — and deserves credit for the courageous and forthright way in which he has denounced the failings of the system and the conservative obstacles to reform.

But Allègre's repeated attacks on the system have become monotonous, and, by provoking a stubborn defensiveness in others, he risks contributing to the very conservatism he has set out to overcome. Moreover, while Captain Allègre's vision of where he wants to go may be clear, winning the support of the crew also demands explaining how exactly he intends to get there. So far, his ministry has failed to do this adequately, and researchers feel insecure as a result. Allègre must not only clarify his intentions but also involve the scientific community more openly in dialogue on how these can be achieved. Perhaps he would do well to reflect on *la méthode Chevènement*. □

Lewis and Clark, lost in the wilderness?

The abandonment of a privately managed satellite should encourage sharper thinking about such ventures.

The demise of the Lewis and Clark satellite projects, which the US National Aeronautics and Space Administration (NASA) had hoped would help to usher in a new era of commercial Earth observation from space (see page 5), could be seen as a mark against the space agency's 'better, cheaper, faster' way of doing business. But that interpretation is too glib. Large, expensive satellites fail, too, and other low-cost projects — such as the Lunar Prospector now circling the Moon — appear to be doing well.

The post-mortem reports on the two satellites should help to determine whether more time or money would have improved either mission's chances of survival. While waiting for these reports, however, NASA needs to examine its often confused policy toward 'commercialization' — the introduction of market forces into the management of space missions.

Lewis and Clark were far from being market-driven ventures. They were NASA's idea, developed with NASA funding. At the time they were conceived, a genuinely private remote-sensing industry had just come into being, and some start-up companies complained that government involvement was unnecessary. Four years later, the fledgling industry has grown to include major companies, such as Lockheed-Martin and Kodak. But its prospects are still uncertain. The first remote-sensing satellite built with private money — the EarlyBird 1, owned by Col-

orado-based Earthwatch, Inc. — lost radio contact with the ground shortly after its launch in December.

Part of the Lewis and Clark experiment was to turn over most day-to-day management decisions to the satellites' private constructors. But this hands-off policy appears to have backfired for NASA. Neither contractor — both experienced companies — succeeded, perhaps because the missions were underfinanced in the first place.

Despite such failure, NASA leaders and many in Congress continue to insist that the private sector can work miracles, and that commercial projects — or even quasi-commercial ones — are by definition better than those run by the government. Commercial space advocates in Congress have been pushing NASA to rely more on Earth science data furnished by the private sector, even though those data do not yet exist. And the agency has reserved substantial room on the space station for customers who may not, in the end, appear.

As they venture into the wilderness of commercial activity, space policy-makers should take care not to imagine markets into being. Outside advice will be most welcome: to its credit, NASA has recently organized a panel of Wall Street advisers to help it to devise a strategy for commercialization, and to sort wishful thinking from cold economic analysis. A small step for the agency; a big step for the future of space missions. □



Max Planck Society faces 'hard decisions' to uphold standards

MPG/FILSER

[GÖTTINGEN] As Germany's Max Planck Society (MPS) turned 50 years old last week, its president, Hubert Markl, told those at the celebration assembly that new measures are essential to prevent the society from drifting from the forefront of research excellence.

The MPS has been home to 15 Nobel prize-winners since 1948, when it succeeded the Kaiser Wilhelm Society. The latter body had also nurtured 15 Nobel prize-winners, but its reputation was irrevocably tarnished by collaboration with the Nazi regime.

The MPS is widely believed to have been so successful because it has been able to offer generous guaranteed funding and excellent research facilities to attract the best scientists as directors of its institutes. But now, said Markl, "competition for the best researchers is growing ever harder". Top scientists are starting to turn down MPS offers, not only in eastern Germany, where living conditions are less comfortable, but also in the west.

Last year, the society decided to dig into its investments to help start an 'excellence in science' fund to improve the employment offers it can make to top scientists. Markl hopes this can be supported by a general improvement in financial conditions. "It would be a waste if we could only attract second-best scientists for this investment," he said.

The society has had to stretch its budget to pay for building up a research base in eastern Germany in the last few years, with only token



Markl: less money will mean "more frequent closures of departments and institutes".

financial support from the government. Parliament blocked a promised 5 per cent increase in the 1998 budget for the society last year, against the wishes of the government.

A positive signal came from Gerhard Schröder, the newly selected Social Democrat candidate for chancellor after next September's federal elections, albeit in the heat of election fever. In his opening address to the MPS assembly last week, he said he strongly regretted parliament's action, and backed an annual increase of 5 per cent for the society.

According to Markl, less money means that there will be more frequent closures of departments and institutes, if the society as a

whole is to respond to new scientific trends. With this in mind, the MPS is embarking on a major exercise called Max Planck 2000+, to identify future research directions for the society's 80 institutes. "This will help make the hard decisions [about closures] which the MPS has never really had to confront before."

Markl also aired his worries about the future of young scientists in the society. It is not clear, he said, whether the society's 2,500 PhD students and 3,000 postdoctoral fellows and visiting scientists are being given the best training.

Markl suggested that the MPS and universities should pool resources to create international graduate schools — an idea that universities are discussing among themselves. He said the MPS is ready "with all its resources" to take part in such developments and to offer "the highest level of international graduate school in defined scientific areas". Courses at such schools should be given in English, he added.

A second idea championed by Markl to promote the career opportunities of young scientists is the so-called '*Nachwuchsgruppe*' programme. This supports small independent research groups, each headed by a talented young researcher, for up to five years. The programme is well-tried and tested in eastern Germany.

The day before the assembly, the MPS signed an agreement with the Weizmann Institute in Rehovot, Israel, to exchange such young scientist groups. Markl hopes to expand this idea to other European countries, such as France.

The MPS has had a cooperation agreement with the Weizmann Institute since 1960. The agreement was partly prompted by the society's desire to atone for Germany's past treatment of Jews. Markl told the meeting that the role of the Kaiser Wilhelm Society in furthering the racist aims of the Nazi regime has been kept under wraps too long.

He also added that the MPS did not get off to an entirely blameless political start. "It is not clear how intensely and systematically the young society tried to call back former Kaiser Wilhelm Society scientists who had been driven abroad [by the Nazis]." Markl assured the assembly of his commitment to openness about those times.

Last year the MPS appointed a committee to investigate the early history of the society. "We can't bask in the reflected glory of the first decade of the Kaiser Wilhelm Society, and not accept our relationship with the dark shadows which fell on science during the Third Reich," said Markl.

Alison Abbott

UK plays safe on risks from blood products

[LONDON] The British government has advised against the use of UK plasma in blood products, to protect patients from the "theoretical risk" of contracting the new variant of Creutzfeldt-Jakob disease (vCJD).

On the advice of the UK Committee on Safety of Medicines, the government is to review all blood products containing UK plasma. Meanwhile the National Health Service Bio Products Laboratory, part of the National Blood Service, will be allowed to import plasma from abroad.

UK plasma, derived from between 20,000 and 66,000 donations each year, is used to make 33 licensed blood products, including coagulation factors and immunoglobulins. Some of these products could have their licences withdrawn, as long as satisfactory alternatives are available, according to Jeremy Metters, deputy chief medical officer.

The committee has also suggested that blood products should be recalled if donors are "strongly suspected" of having vCJD.

Previous product recalls were based only on confirmed cases. The latest move follows three recalls of blood products last November after donors had developed vCJD.

Frank Dobson, the health secretary, stressed that the recommended measures are strictly "precautionary" and that there is no evidence of vCJD transmission through blood or blood products. But he said it was better to be safe than sorry. "If there is even a hypothetical risk, and there are safe alternative sources of products, then it makes sense to use them."

Dobson also announced the outcome of a review of the NHS's provision of factor VIII. He said health authorities are being urged to ensure that recombinant factor VIII is made available to haemophiliacs under the age of 16, and to new patients. The review highlighted concerns over the potential danger of using factor VIII, as haemophiliacs have a high potential risk of contracting blood-borne infections.

Asako Saegusa

Russia invests in a 'revolutionary' disc

[MOSCOW] A Russian-born physicist who emigrated from the Soviet Union in 1975 has persuaded the government to invest in a factory to make what he claims is a "revolutionary" three-dimensional compact disc, able to store information at up to 50 separate levels.

Eugene Levich has recruited Egor Stroev, speaker of the Federation Council — the upper chamber of the Russian parliament — as chairman of the company, 3-DOM, set up to produce the discs. When Levich left the Soviet Union in the 1970s he went with his father, Veniamin Levich, a corresponding member of the Academy of Sciences.

Sergey Shcheblygin, one of Stroev's advisers, confirms that the Russian government has deposited money in an account, confirming its intention to participate in the project. Shcheblygin says 3-DOM is not entitled to spend the money, but Levich insists the company can use it as it wishes, a view supported by fellow director Iosif Diskin, deputy director of the population institute of the Russian Academy of Sciences.

3-DOM will make multilayer fluorescent discs (MFDs), which it claims can store tens or hundreds of times more information than standard discs, but at the same price. According to Diskin, the design of the discs is based on so-called 'stable photochrome', discovered by physicists and engineers who emigrated from the former Soviet Union.

This is a transparent organic substance which fluoresces when subject to a laser beam for sufficient time to be detected by a standard photo-receiver. This characteristic makes it possible to superimpose up to 50 transparent layers on top of one another, and to 'write' information on each level, which can be read either in succession or in parallel.



Disc man: a researcher at Moscow State University Science Park with the 3D compact disc (right).

The science has been known about for some time in the West, and is said to have been investigated by companies such as IBM and Philips, but the reading-devices needed are so complex that none has made any public commitment to invest in the technology.

"The scientific aspects are sound, but it is still untried as a technology," says Nicolaas Bloembergen of Harvard University, 1981 Nobel prize-winner in physics, who has been recruited to 3-DOM's scientific advisory board. "It appears to work in a laboratory demonstration. The big question is whether it is worth taking the technological risk."

Diskin says production of the discs is due to start in Russia by next summer. But Levich admits that so far all MFDs have been made by hand, and that a production line has still to be designed. Both have told Shcheblygin that specifications are being drawn up for equipment to be installed in the former Control electronic machines factory in Orel, but decline to provide further details.

Diskin says the initial plan is to use MFD for high-definition television, as current DVD discs cannot store sufficient information for a two-hour video. "Our technology is so simple that producing a disc holding a high-definition TV film worth \$100 costs us practically nothing," he says. He predicts that in one year, when the factory starts production, it could pay taxes comparable to "the whole budget of Orel district".

The financial backing to the project is complex. The Austrian-registered company Constellation Group owns 20 per cent of the shares in 3-DOM, as well as the capital of two other companies, the US-based Constellation Memory Division Inc., (which owns 20 per cent of 3-DOM shares) and High Technology Projects (14 per cent). Levich heads the first two companies, and Diskin the third.

The remaining capital is held by the Russian district administration of Orel, of which Stroev is governor (33 per cent); the state property foundation (10 per cent); and Moscow State University (3 per cent).

The university's participation is through its Laser Centre, which signed a contract in 1995 with Memory Devices of Maryland, another company created by Levich and Diskin. Under this agreement, the laser centre has developed a two-photon, single beam 'bit-by-bit' writing technique, and a fluorescent 'page-by-page' reading optical data storage system for use with MFDs.

Levich says that the émigré physicists had a choice of offering the technology to the Lawrence Livermore Laboratory in the United States or the International Laser Centre at Moscow State University.

"The fact that two-thirds of Constellation Group are Russians, and that the Russian workforce is less expensive, were not the main reasons we chose Russia," says Levich. "In the United States, people would not listen to our 'crazy' ideas, and kept asking how we had obtained these results while IBM or Sony had failed."

3-DOM's science advisory council is chaired by Viktor Sadovnichy, rector of Moscow State University, and includes prominent scientists such as John Armstrong, formerly chief scientist of IBM, and Mikhail Alfimov, chairman of the Russian foundation for fundamental research.

But Bloembergen says this committee has never met. Promises by Levich that the committee would meet in January, and that samples of the disc would be made available last month, have also failed to materialize. And neither Diskin nor Levich will reveal how to contact the Constellation Group, or any of the companies they have created.

"We are a small group and need to be very cautious surrounded by many powerful rivals," says Levich.

Carl Levitin

India urged to act against leptospirosis

[NEW DELHI] Medical researchers in India have urged the government to take prompt action to prevent leptospirosis from becoming a major public health problem following several outbreaks of the disease, which is spread mostly by rats.

Last year, 90 people died in four outbreaks, including one in a medical college hostel in Mysore whose drinking water was contaminated with rat urine. Health officials fear it may be impossible to control the disease unless an effective rat control strategy is found. Little is known about the identity of other potential animal carriers of leptospirosis.

Officials point out that the disease can be treated with antibiotics such as penicillin. But Vulimiri Ramalingaswami, former chief of the Indian Council of Medical Research,

says it is essential to develop tools to detect leptospirosis and to launch a surveillance programme. "Otherwise, India would face a major public health problem in the event of [the disease-causing bacteria] developing resistance to penicillin."

Leptospirosis causes a high temperature and jaundice, and affects the lungs, kidneys and brain. It is caused by a spiral bacterium, *Leptospira interrogans*. Infected animals shed the bacteria in urine and people coming into contact with contaminated water or soil pick up the infection through the skin and mucous membrane.

Experts told a conference in New Delhi last week that mushrooming slums in cities and the rise in development activity in villages provide ideal environments for the disease to spread.

K. S. Jayaraman

Clark's demise halts NASA experiment in remote sensing

[WASHINGTON] The second of two Earth-observing satellite projects designed to help jump-start a commercial remote sensing industry has been scrapped by the US space agency NASA after cost and schedule overruns. The cancellation of the Clark project follows the failure of its companion satellite, Lewis, shortly after its launch in August (see *Nature* 389, 108; 1997).

The two spacecraft were intended to develop techniques and technologies that would be useful to future Earth-observing satellites, including hyperspectral imaging and the ability to selectively 'edit' out undesirable images obscured by cloud cover to conserve data.

NASA has long known that Clark was in trouble. The project was announced in June 1994 and was due to launch in March 1996, but difficulties with some of the spacecraft's subsystems and delays associated with a new rocket built by Lockheed Martin Aerospace pushed the date back by more than two years.

While Clark was being developed, the company originally contracted to build it, CTA Inc. of Rockville, Maryland, was purchased by Orbital Sciences Corp. of Dulles, Virginia, which also makes the Pegasus rocket. NASA had hoped that Orbital would solve Clark's management woes, but the cost overruns and technical problems continued.

After a review team at NASA's Goddard Space Flight Center recommended cancellation earlier this year, agency administrator Daniel Goldin acted on repeated threats to terminate any agency development project that runs more than 15 per cent over budget.

The space agency spent about \$55 million on Clark, but not all of that is lost. NASA will retain launch-vehicle services from Lockheed Martin, and some of the satellite subsystems can be used on future projects. Developers of the software algorithms for cloud editing expect to use the technique in other missions.

A report on the failure of the \$71 million Lewis satellite is due from NASA this month. A panel chaired by Christine Anderson, director for space vehicles at the Air Force Research Laboratory in Albuquerque, New Mexico, investigated the incident and has already presented its findings to NASA.

TRW Inc., which built and operated Lewis, has completed an internal investigation, but is not commenting publicly on the reasons for the failure. NASA sources have said that deficiencies in tracking and monitoring the satellite's progress after launch were partly to blame. Lewis might have been saved, they say, if spacecraft operators been present in the control room during the period of several hours when the spacecraft first began to spin out of control. **Tony Reichhardt**

Survey results boost calls for new teaching efforts

[WASHINGTON] An international study showing that 17-year-olds in the United States are performing poorly in mathematics and science could serve as a wake-up call for action to raise school standards, according to US scientific leaders.

"This is a great opportunity to mobilize scientists and engineers," says Bruce Alberts, president of the National Academy of Sciences, referring to the latest results from the Third International Mathematics and Science Study (TIMSS). But Alberts and other scientific leaders stress that any such response will need to be undertaken voluntarily at a grass-roots level, and be directed at each of the 16,000 autonomous district authorities that control US schools.

Unlike the governments of Sweden, Norway, Denmark and most of the other states that comfortably out-performed the United States in the study, the US federal government lacks the power to take direct action to reform science education in schools. "Education has been taken away from federal control," Alberts says. "We're not going to fight that battle again."

The study, which compared the performance in mathematics and science of 17-year-olds in 21 countries, as well as the performance of advanced students in these subjects in 16 of the countries, placed US students in the lowest band in every category.

The United States ranked third-last in mathematics — above Cyprus and South Africa — and sixth-last in science. The performance of its advanced students was even weaker: they came second-last in mathematics and last in science. The study did not include most Asian nations or the United Kingdom, so the comparison was chiefly with European nations, Canada and South Africa.

President Bill Clinton responded to the results by calling for more professional development for teachers and more widespread acceptance of national standards in schools. "I hope that eventually we will have every state testing their children in basics and measuring them by a common national standard," he said.

Neal Lane, director of the National Science Foundation, took comfort from the fact that the US has fared better in a 1995 TIMSS survey of nine-year-olds. Lane said that younger children were benefiting from recent government initiatives and predicted that "we can expect to see improvement in their performance at higher grade levels" in future. TIMSS will report a new comparison of 13-year-olds next year.

But few observers share Lane's belief that the United States will have improved by then.



Local knowledge: Richard W. Riley, Secretary of Education, visits a Baltimore high school.

Some science educators see the TIMSS study as the inevitable result of a fragmented education system, in which attempts by the federal government to impose standards on schools has opened up deep political divisions.

Clinton has proposed various initiatives to expand the role of the Department of Education in schools, but the Republican Congress values the right of states and local school boards to set their own education policies. Republicans have strongly opposed national testing of students, and were until recently seeking to abolish the department.

"We are very concerned about the TIMSS results," says Bill McCarthy, a spokesman for the Education and Workforce Committee in the House of Representatives. "But we still believe that education is best handled as a local issue."

This desire for autonomy is at its strongest where it affects the teaching of history and other politically charged topics. But the teaching of Darwinian evolution has also generated frequent controversy. Vernon Ehlers (Republican, Michigan), the vice-chairman of the Science Committee in the House of Representatives, says that evolution is "probably the only example" of such contention in science, and argues that Republicans will support the use of national science teaching standards by school districts.

Such standards have been developed and published by the National Academy of Sciences. Alberts says that, although the effort was publicly funded, "it has been accepted as a national, not a federal, standard". But the decision to use the standards rests with each of the thousands of school boards. So when the federal government is unable to intervene, what is going to influence their autonomous actions? "Maybe these test results will," says Alberts. **Colin Macilwain**

Privacy bill under fire from researchers

[WASHINGTON] An effort to write a comprehensive law to protect medical privacy was launched in the US Congress last week. The initiative is supported by privacy advocates, but has been criticized by researchers and industry lobbyists, who argue that it would create damaging barriers to the use of patient information in research.

At the centre of the debate is the Medical Information Protection Act of 1998, a draft bill from Senators Robert Bennett (Republican, Utah) and James Jeffords (Republican, Vermont), chairman of the Senate Labor and Human Resources Committee. It was presented at a committee hearing last week.

The bill would establish protection against — and criminal and civil penalties for — the improper disclosure of protected patient information. Most of these provisions would affect the use of information by the broader health-care industry.

The bill extends to the private sector federal protective measures for human research subjects that now apply only to government-funded researchers. These would require pharmaceutical, biotechnology and health-care companies to obtain informed consent from subjects, and approval from local ethics committees called institutional review boards (IRBs), for a wide range of research using identifiable records that currently goes forward without federal oversight.

Of greatest potential concern to biomedical researchers — most of whom already operate under federal rules on the use of human subjects — are provisions that would affect scientists who rely on the use of existing, 'identifiable' medical records, or the estimated 283 million archived, 'identifiable' pathological specimens.

Such records and specimens carry either patient names or codes that would ultimately allow patient identification. Two paths generally provide ready access to these resources.

First, under federal regulations, an IRB



can waive the informed consent requirement for the use of identifiable records or specimens in retrospective studies, if it accepts that it would be impractical to do the research if informed consent had to be obtained, and that the study would pose no more than minimal risk to subjects. Alternatively, researchers seeking access to identifiable records or specimens can use 'expedited' IRB review, in which one board member is authorized to rapidly approve such studies.

The Bennett-Jeffords bill would eliminate expedited review, forcing formal IRB review for such studies. It would also bring IRB waivers of informed consent in these cases under much closer scrutiny. For instance, in what critics complain is unduly vague and subjective language, the board would have to determine that a waiver is "necessary for the effectiveness of the research".

The bill also requires the Institute of Medicine to study the "research issues relating to protected health information, such as the quality and uniformity of [IRBs] and their practices with respect to data management for researchers".

Lobbyists for medical researchers say the draft bill is a threat to valuable research that has only been made possible by the

flexibility of the current regulations.

"There is real concern that the IRBs will be inundated with this huge amount of research that they have never had to worry about," says David Korn, vice-president for biomedical and health sciences research at the Association of American Medical Colleges, which is drafting a critique of the bill.

"The concern is that they will be tied up in knots trying to debate within their own members what to require of this research," says Korn, a pathologist by training and the former dean of the Stanford University Medical School.

The Pharmaceutical Research and Manufacturers of America last week issued an extensive, critical analysis of the bill. It complained that the IRB system would be "quickly overburdened" by an extension of the requirement for approval to non-government-funded research.

But privacy advocates say the time is ripe both for that extension, and for more scrutiny of how IRBs deal with 'identifiable' records. Research based on records, they say, can put human subjects at as much risk as participating in trials.

IRBs are "more comfortable and more used to focusing on people than information," says Janlori Goldman, director of the Health Privacy Project at the Institute for Health Care Research and Policy at Georgetown University in Washington DC, who testified at the Senate hearing. "But this is a cultural change that must occur within the research community."

Some researchers agree. Ronald Levy, chief of the division of oncology at Stanford University Medical School, has relied on archived specimens and records for his work on non-Hodgkin's lymphoma. While he is comfortable with the status quo, Levy says: "I don't know how one can make an argument that there should be an easy way to violate privacy," although describing the potential loss of expedited review as "a nuisance".

The bill was to have been introduced last week, but was delayed because of criticism from privacy advocates that it did not clearly block the release of patient data to marketers of medical products. This issue is high on the public agenda in Washington since it was reported recently that the drugstore chain CVS and Giant Food Inc., a grocery-pharmacy chain, were selling prescription information to a marketing company.

An official on Jeffords's staff says the bill will be introduced within a couple of weeks. But lobbyists say this is unlikely, given what they call the bill's extensive problems.

Under the Health Insurance Portability and Accountability Act, which became law in 1996, Congress must write privacy legislation by August 1999.

Meredith Wadman

Biomedical billions go under the microscope

[WASHINGTON] An inquiry into how the US National Institutes of Health (NIH) sets priorities for research spending will begin this week with a public meeting in Washington. The study, by the Institute of Medicine (IOM), was ordered by the US Congress in the law passed last year to fund the NIH in 1998.

The aim of the \$338,000 study is to recommend improvements in research

funding policies and to suggest any necessary congressional action.

At this week's meeting, NIH officials will explain to a committee from the IOM how they divide the funds in their \$13.65 billion budget. They will also describe how they listen and respond to public opinion — which is volatile, with the constituencies for different disease groups arguing for more funds.

Officials will tell the

committee how Congress affects the agency when it passes laws requiring NIH to spend money on research on specific ailments, from Parkinson's disease to diabetes.

The next meeting, on 3 April, will receive public comments on NIH's priority-setting. Comments can be posted electronically to the committee's Web site at <http://www2.nas.edu/hsp/214e.html>.

M.W.

Cuts reversed in Canada's new budget

[MONTREAL] The most extensive lobbying ever conducted by Canada's research scientists has resulted in a substantial increase in federal funding for the three major grants councils — the first increase in three years.

After a period of steady decline, the councils' overall funding will rise in one step by 14 per cent over this year's figures. The increases appear in the country's first balanced budget in 29 years, presented last week by the governing Liberal party.

The Medical Research Council's (MRC's) budget for 1998–99 is to rise from C\$238 million (US\$167 million) to C\$267 million, rather than falling as planned to C\$227 million. The figures for the following two years are C\$270 million instead of C\$226 million for 1999–2000, and C\$276 million instead of C\$226 million for 2000–2001.

The Natural Sciences and Engineering Research Council budget will receive an even greater rise, to C\$494 million instead of C\$423 million in 1998–99, C\$495 million instead of C\$423 million the following year and \$501 million instead of C\$416 million the year after that. The current budget is C\$434 million.

The Social Sciences and Humanities Research Council budget rises will be C\$101 million, C\$101 million and C\$103 million.

The differences in rate of increase between the three councils result from a decision to give each the same share of the overall research budget this year as it received in 1995, according to John Manley, the industry minister.

An important factor in the success of the lobbying was a mass campaign by bench scientists of letter-writing and visits to members of parliament and ministers.

The increases have been widely welcomed by the scientific and academic communities. But some point out that the government is just restoring the grants council budgets to their 1994–95 levels, without taking inflation into account. Others complain at the enormous disparity that remains in per capita funding between Canada's MRC and the US National Institutes of Health (NIH) — a ratio of about 1:8 (see *Nature* 390, 210; 1997). The NIH competes in the same areas of research with grants on average one-third higher, making it easier for the NIH to attract top scientists.

Despite such reservations, Canadian medical science benefited immediately from the budget changes. After the last meeting of its peer-review panel in September, the MRC was forced to cut by 26 per cent the number of projects it had approved because of lack of

funds. Two days after the budget was announced, the council confirmed that 109 of these projects would be funded, and extended another 26 from one to three years. Funding was also announced for a postponed C\$500 million clinical trial.

The budget also includes the creation of a C\$2.5 billion Canada Millennium Scholarship Foundation to provide more than 100,000 scholarships a year to low- and middle-income students. This was denounced by Quebec separatists as an intrusion into provincial jurisdiction, but praised by others.

New tax measures will help post-secondary students. An extra C\$205 million over three years will expand a programme to provide access to the Internet by schoolchildren and others.

Climate-change projects will receive C\$50 million a year for three years, plus C\$34 million a year through the National Research Council's Industrial Research Assistance Program. The latter is to help businesses foster technologies for using energy and natural resources more efficiently and preventing pollution. But the relatively small size of this contribution to Canada's commitment at the Kyoto climate conference was derided by Jim Fulton of the Suzuki Foundation, an environmental action group. **David Spurgeon**

Australia urged to avert collapse of university science

[SYDNEY] Australian universities are calling on the government to rescue academic science from what John Niland, the new president of the Australian Vice-Chancellors' Committee, describes as an impending "national emergency akin to a natural disaster". This marks the first time the 36 member universities have united in support of a single field of study and research.

Niland, a professor of industrial relations and vice-chancellor of the University of New South Wales, delivered the challenge in Canberra last week in a televised address to a forum of academic and business leaders organized by the Federation of Australian Scientific and Technological Societies (FASTS).

John Rice of Flinders University, president of the Australian Council of Deans of Science, says "university science is in danger of spiralling out of control". He says the government is making "fundamental errors in believing universities are tertiary high schools whose sole customers are undergraduate students, and that research funding can be picked up by the private sector".

Niland says the "deep trouble" facing science is being felt most in "the very foundations of science" — physics,



Niland: blames crisis on government cuts.

mathematics and chemistry — where student interest is low, rather than in computer science or the life sciences where enrolments have grown.

Niland says a public opinion survey identified the following factors as contributing

to the low morale in science:

- a sharp increase in tuition fees — from A\$2,250 (US\$1,512) in 1996 to A\$4,779 (US\$3,211) in 1997;
- fewer teaching staff — student/staff ratio rose from 12.3 in 1988 to 16.7 in 1997;
- ageing facilities (universities receive only about 70 per cent of the funding of those in the United Kingdom and Canada); and
- difficulties of beginning a research career as grants become harder to obtain.

According to Niland, a pass degree in accounting or law can lead to twice the salary of a professorship in science. An international index of salaries in research and development shows that, for every A\$100 paid to the average Australian scientist, the United Kingdom pays A\$104, France A\$116,

the United States A\$141, Japan A\$158, Hong Kong \$166 and Germany A\$170.

Niland claims "the crisis for science" is a result of government cuts to the university sector. He calculates from the government's current budget plans that the cuts will amount to 30 per cent from 1997 to 2001 if the anticipated growth in national productivity is included.

The government argues that the research situation in universities is far from critical. In a statement to the forum, education minister David Kemp wrote of "modest reductions" in the 1996 budget which mean that operating grants "will fall by less than one per cent between 1996 and 2000". Kemp has not responded to the campaign, which is pressing him to restore science funding to international levels in the May budget. The government plans to use the budget to return public finances from deficit to surplus.

Others say that universities need to develop some self-criticism, in addition to targeting the government. Peter Cullen, president of FASTS, says universities must not only press for better funding: "They need to be rattled by this crisis into reshaping curricula and organization into a more effective balance between the basic and applied sciences." **Peter Pockley**

Jury still undecided on French reforms

Declan Butler

Hopes were high within the French research community when Claude Allègre was appointed minister of research last year. But a good start has given way to frustration with the slow speed of change.

[PARIS] Claude Allègre, the French minister of national education, research and technology, announced last week that the government is to organize an interministerial council in April, chaired by Prime Minister Lionel Jospin, to "propose a structure and priorities for French research".

Allègre said that while a detailed agenda remained "confidential", the outcome would be a "republican programme". This would link research to national goals — for example educational and industrial development and the prevention of natural catastrophes — and orientate biomedical research so as to reduce the health-care deficit.

Some observers hope the meeting will also serve to clarify Allègre's intentions for reforming the research system. When Allègre was made minister after last June's election of a socialist government, he promised a profound shake-up of the country's research system.

Nine months later, many in the research community are impatient at the slow rate of progress, and increasingly worried that Allègre's bruising, single-minded style is itself an obstacle to change.

"For the moment I have seen nothing [in terms of reforms] that will help solve the problems of French research," complains Pierre Chambon, head of the Institute of Genetics and Molecular Biology near Strasbourg. He deplores "the absence of any innovative or substantial initiative. Nine months is the length of a pregnancy, but we have yet to see the baby." Other researchers, who wished not to be identified, expressed similar feelings.

Many of these argue that, after the election, conditions were ideal for Allègre to embark on radical reform. He won widespread confidence and goodwill in the scientific community by creating thousands of new posts for researchers and technicians. He also negotiated a modest increase in science spending when France was cutting public spending to reduce its deficit to meet the criteria for joining the Euro.

Allègre himself brought excellent credentials, including his outspoken support for fundamental research and his energetic track record as Jospin's principal adviser when the latter was education minister in 1988–92. A close friend of the prime minister, Allègre also obtained a superministry, giving him unprecedented power over all areas of gov-



Allègre's intentions for change are still unclear to an impatient research community.

ernment science and technology.

But the perceived lack of substantial progress towards reform is creating frustration among many in the research community. "All the elements were in place for the scientific community to be optimistic, but the practices of this minister means that it is now relatively discontent," admits one official at the Centre National de la Recherche Scientifique (CNRS).

Allègre himself is increasingly being seen as an obstacle to reform. In his short spell in office he has gained a reputation for lack of tact, and for repeated provocative attacks on the education and research system. This resulted in such discontent that Jospin recently reprimanded him publicly. Some consider that Allègre's lack of diplomacy is losing him the confidence and goodwill of scientists, and perhaps with it the opportunity for cooperation in reform. "I think Allègre's belief is that reforms cannot be achieved by consensus and that you have to be aggressive," says one senior research official. "But I find that it is stalling the system, as people are losing confidence; the ministry's method has become the principal ally of inactivity."

Writing in the newspaper *Le Figaro* last week, Allègre took a more conciliatory tone. He argued that reforms should not disturb what already works, and that improvements should be made incrementally. Several scien-

tists say they would not object to Allègre's aggressive tactics were these necessary to overcome conservative resistance to worthwhile reforms. But they argue that the reforms he has introduced have been of such limited value that his aggressive strategy has been unwarranted and counterproductive.

Allègre's proposed reforms of the CNRS national committee have been criticized, for example. The committee, which consists of around 40 sections covering various disciplines, comprised of scientists from the research agencies and universities, is responsible for evaluating all CNRS laboratories and administering recruitment to CNRS; in practice it also has a major influence in shaping CNRS's research priorities.

Allègre has repeatedly stated his desire to see the number of committees halved as part of his aim to reduce bureaucracy. He argues that too many scientists spend too much time in committees instead of on research. But this proposal, which has been fiercely resisted by the scientific trade unions as a "declaration of war", has been roundly rejected in a report commissioned by CNRS from Jean Pailhous, a CNRS researcher at the Université Aix-Marseille 2.

The report, which is based on a broad consultation, recommends that the number of committees be only slightly reduced, to 36, and proposes a redefinition of the committees' remit to correspond better with how disciplines have changed. The report makes a point — shared by many scientists — that evaluation requires considerable time and resources, and that a drastically reduced national committee would lack the manpower and expertise to do the job properly.

This conflict is likely to be only the first salvo in a wider debate over the future of CNRS's operation. Henry Edouard Audier, a chemist at the Ecole Polytechnique, says, however, the ministry has accepted his proposal for a three-month consultation on the committee's future.

More broadly, several scientists also argue that the ministry's focus on the number of CNRS committees distracts it from more pressing concerns. "Reducing the number of committees will change nothing; it is not in making reforms like that that we will solve the problems," says Chambon.

Reform at a snail's pace

Observers attribute the perceived lack of progress to several causes. According to some, Allègre appointed too many new faces to senior positions within the research agencies and the ministry, with a resulting loss of experience, competence and continuity.

The sheer size of the jurisdiction of Allègre's ministry is itself hindering progress, according to others; they point out that the day-to-day demands of the schools portfolio alone are enormous. Responsibility for research reforms has fallen largely to Vincent

Courtillot, Allègre's principal adviser.

Several senior research officials argue that while Allègre is personally flexible and open to discussion, his cabinet is "rigid" and has shown a reluctance to consult with others in the research system. "In practice the cabinet is more difficult to work with than Allègre," says one CNRS official. He added that working relationships now seem to be improving and that the cabinet is "more willing to listen". Efforts to contact Courtillot for comment last week were unsuccessful.

Scientific community open to change

Frustration over the lack of visible progress on reforms is even greater because, whereas the research community has often resisted reform, the government's strong support for science has created a climate conducive to change, and the need for reform is now widely acknowledged. Indeed, judging by a major conference on biomedical research policy organized by parliamentarians and scientists in Paris last week (see below), Allègre's promises of deep reforms seem to have created a thirst for change in the community.

A wider reform of the evaluation system is considered a priority by many scientists. They believe that a major challenge for French research is to distribute funds and resources more efficiently, making the system more competitive. Many would like to see funds distributed to individual research groups on the basis of competitive grant applications, instead of through the present system in which funds are spread broadly among the public research agencies and university laboratories.

Critics argue in particular that the system favours established scientists, and that change

would increase the competitiveness of French research by providing greater opportunities for young scientists. A more elitist system would also reduce mediocrity and the dispersal of scant resources, they add. Allègre, who has said he considers research groups as the "basic unit of research", is seeking a deeper reform of the evaluation system, involving for the first time international experts.

Catherine Brechignac, the director-general of CNRS, says she is opposed to evaluation on the basis of research groups, and argues that laboratory directors are able to evaluate their own groups. Meanwhile, Brechignac has decided to reduce the frequency with which CNRS laboratories are evaluated from every two years to every four, arguing that this will allow groups to relax and take on more ambitious projects, while reducing bureaucracy.

Increasing role for universities

Another area in dire need of reform is the relationship between CNRS and other research organizations and the universities. There are two competing visions: that CNRS should remain an independent agency, much like Germany's Max Planck institutes, or that there should be a progressive fusion of CNRS laboratories and the universities.

Over the past two decades the latter has increasingly dominated thinking, and many believe that a major leap in this direction is now imminent. Allègre has made no secret of his belief that the universities should be the driving force of French research, and that his model is the Anglo-Saxon research university.

The creation in the 1970s of mixed laboratories between CNRS and universities is widely credited with having transformed the research landscape in France, irrigating the



The old and the new: CNRS's headquarters outside Paris are set for further changes.

university system with laboratories working in many areas of basic science. With the lessening pressure on university budgets from the explosive growth in student numbers during the 1980s, many see the universities as being well placed to lead French research.

Even Edouard Brezin, the president of the CNRS board, questions whether CNRS should continue in its present form. "The major flaw with CNRS at present is that research funds are spent on a staff of 11,500 people employed for life within CNRS," says Brezin. "This is not the best way to use resources."

Brezin says there is now an opportunity to "change the shape of CNRS" by reducing the number of staff employed for life within the agency itself. The remainder of CNRS posts could then be more mobile, staff being recruited only for during most productive years and then moving on to the universities.

Several observers speculate that Allègre would ultimately like the research agencies to be transformed into research councils, with the universities having responsibility for all laboratories. Brechignac says that CNRS's national character makes it better geared than the universities to developing research strategy, and that CNRS laboratories therefore "complement" the university system.

One change Allègre has often promised is a reorganization of France's research agencies. He argues that they have become isolated from one another, and progressively strayed beyond their core missions. The Atomic Energy Commission now has many laboratories in climatology and Earth sciences, for example, while life science research is done there and at CNRS and INSERM.

But the expected radical reshuffle has yet to materialize, with Allègre arguing that inter-agency cooperation must be increased. For example, instead of creating a single agency for biological and medical sciences, as suggested by several scientists, the government has opted to set up a ministerial committee to coordinate such research among the various agencies, an initiative judged as inadequate by many biologists.

This approach is defended by Brechignac, who questions the usefulness of dramatic changes. She argues that efficiency is better achieved by closer cooperation and by modifying the existing agencies.

Researchers challenge biomedical plans

Critics cite the government's recent unsuccessful attempt to impose changes on INSERM, the national biomedical research agency (see *Nature* **391**, 110; 1998) as evidence of the shortcomings of the government's reform efforts. Allègre has criticized the agency, for example, for failing to underpin applications of medical research such as telemedicine, biotechnology and new drugs.

To remedy the situation, he has proposed a greater emphasis on these applied goals and splitting it into distinct departments. But the reforms have been vigorously challenged by many

scientists as being poorly thought through and likely to have little impact, while damaging INSERM's fundamental research capacity. Christo Goridis, head of the joint CNRS/INSERM Institut Fédéraliste de Recherche de Biologie du Développement in Luminy, near Marseilles, argues that splitting the relatively small agency into several departments would create "unacceptable" artificial barriers.

Henry Edouard Audier, a chemist at the Ecole Polytechnique and a member of the board of the national researchers' trade union SNCS, argues that the 'diagnosis' – that France is

weak in these industrial areas – is correct, but that the proposed treatment would have done little to remedy the situation, since the causes are much wider and related more to deficiencies within the industries themselves than within INSERM.

The original reforms have since been rejected by INSERM's representative bodies, while the idea of creating new departments is said to have been watered down to the creation of committees within INSERM's management. Allègre last week threatened that if INSERM does not agree to this, the government will simply take such research "elsewhere". **D.B.**

Pig virus discovery underlines risks of xenotransplants

[SYDNEY] Australian researchers say they have discovered a virus that has 'jumped' species from fruit bats to pigs, and apparently to humans. The finding followed a marked reduction in the size of litters and the birth of many stillborn piglets on three farms in New South Wales. Many of the piglets surviving suffered birth defects, notably in the brain and spinal cord.

Peter Kirkland, of the Elizabeth Macarthur Agricultural Institute near Sydney, told the Horizons of Science Forum in Australia last week: "This had not been seen in pigs before." Kirkland classifies the agent as a paramyxovirus, a group including canine distemper and human measles.

Tests on cattle, sheep, cats, birds, rodents and a dog near the pigs were negative, but finally the virus was traced to a neighbouring colony of fruit bats. Antibodies to the virus were found in two workers out of 36 on the farms who contracted influenza-like illness at the time the virus was affecting the pigs.

Kirkland does not believe the virus presents a health risk to the public, as it is not highly infectious. But the discovery has highlighted the potential hazards of using tissues from pigs for transplantation to humans. Kirkland says that screening programmes for testing donor pigs could not be expected to successfully detect an agent such as this "especially when it is transmitted silently in the pig population."

Court ruling could close medical marijuana clubs

[WASHINGTON] The Supreme Court of California last week unanimously refused to review a lower court decision prohibiting 20 medical marijuana clubs from selling the substance. The clubs may now be shut down.

The clubs claimed that they should be allowed to supply the drug under the terms of a 1996 referendum in which California voters approved the use of marijuana for medical conditions such as glaucoma, cancer and AIDS, allowing its possession and cultivation for these purposes on a doctor's orders.

But last December the First District Court of Appeal ruled that the referendum did not allow clubs to provide marijuana to patients who designated a club as their 'primary care giver'.

Germany and US agree energy research pact

[MUNICH] Germany and the United States have signed their first energy research cooperation agreement. One of the goals of the agreement will be to find ways to fulfil

targets set at the Kyoto conference on climate change, for example through the development of renewable energy sources.

The terms of the agreement anticipate specific technology partnerships. These could include a two-year project to develop a new generation of superconducting magnets for high-energy accelerators, involving DESY, the German electron-synchrotron facility, and the US Brookhaven National Laboratory.

Radioactive pollution at sea sparks Nordic protest

[LONDON] Environment ministers from Nordic countries are calling on the United Kingdom to stop releasing into the sea the radioactive substance Technetium-99 from the nuclear reprocessing plant at Sellafield in the northwest of England.

The call was made after a meeting of the Nordic council of ministers last week. It was prompted by the detection of raised concentrations of Technetium-99 in seawater along the coast of Norway during 1996 and 1997.

Anna Lindh, Sweden's environment minister, said: "The Nordic countries with their extensive coastlines are particularly vulnerable to pollution of the marine environment. Many of the countries are very dependent on fisheries and clean fishing waters."

'Risky' research wins more British backing

[LONDON] The UK government is to extend a £100 million (US\$165 million) scheme designed to fund riskier strategic research of potential help to industry. The Realizing Our Potential Award (Ropa) scheme will now award funds across all scientific disciplines covered by all the research councils.

The research councils have given 1,200 Ropa awards since the scheme was set up in 1994. A Ropa award is considered successful if it generates a research advance that justifies further exploration.

Car makers in a jam over emission targets

[LONDON] Car makers are locked in a dispute with Europe's governments and the European Parliament over reductions to carbon dioxide emissions from vehicles.

Manufacturers have offered to cut emissions to an average 155 grams of carbon dioxide per kilometre in new cars by 2005, compared to the average in 1995 of 171 grams. But the governments are insisting that the companies agree on a target of 120 grams to help achieve the European Union's overall legally binding reduction targets agreed at the United Nations climate

conference in Kyoto last year. Motor manufacturers say they fear the union's proposed targets could drive them out of business. But the European Commission is understood to be preparing to draw up mandatory targets if the companies fail to agree on a voluntary cut.

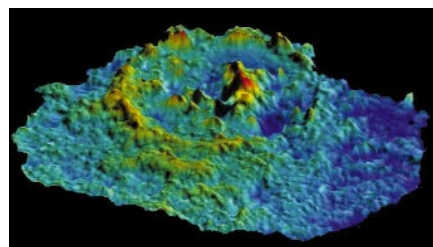
£3m scheme promotes biotech business skills

[LONDON] Ten projects have been picked for a £3 million (US\$4.5 million) UK government initiative to provide entrepreneurial and intellectual property management skills to bioscientists.

Five of the projects come under the Biotechnology Exploitation Platform scheme. This aims to build partnerships between research groups and experts in marketing and intellectual property rights management. The other five awards are part of the Biotechnology Mentoring and Incubator Challenge, designed to help small biotechnology companies start up and grow.

The winners include Oncotech, a consortium made up of the Cancer Research Campaign technology transfer company, the Leukaemia Research Fund, the law firm Cameron McKenna and the patent firm Mewburn Ellis. The group's aim is to set up a company to focus on maximizing the commercial potential of cancer research.

NASA gives a new angle on jovian moon's water



[WASHINGTON] Pwyll Crater on Jupiter's moon Europa (above) shows some of the best evidence yet for water or warm slush underlying the moon's crust of hard ice. In this computer-generated view based on data returned by the Galileo spacecraft in February and December 1997, and released by NASA on Monday (2 March), the floor of Pwyll is seen to be shallow, lying at about the same elevation as the surrounding terrain.

The crater's central peak, which stands 600 metres above the floor, is higher than its rim, suggesting that Pwyll was modified by slushy ice after its formation by impact between 10 million and 100 million years ago. The vertical scale of the image has been exaggerated. The space agency was expected to make an announcement today (5 March) about its most recent conclusions concerning the possible presence of water on the Earth's Moon.

Xenotransplants: proceed with caution

Sir — The US Public Health Service (PHS) agencies have rejected recommendations for a moratorium on clinical trials of xenotransplants, and left the door open for nonhuman primates to be donors. Your Briefing on xenotransplantation¹ summarizes the important elements of this subject, and captures perfectly the split between those who want to get it right and those who want to get it right now.

In the Briefing, Jonathan Allan, a virologist on the Food and Drug Administration (FDA)'s advisory subcommittee on xenotransplantation, points out the neglect of the precautionary principle in a situation where it is most needed, given the risk of public exposure to xenozoonoses. The United States is fortunate in having bodies such as the Centers for Disease Control and Prevention, the National Institutes of Health and the FDA, which give it the confidence to halt activities "should something happen" — but containment is the wrong strategy here, and any epidemic that starts may involve countries other than the United States.

Many countries, and not just those in the developed world, are likely to want to benefit from xenotransplantation. Given that nothing is to stop them from going ahead, apart from their national governments — if even that — the need for international cooperation in risk minimization and containment becomes obvious.

Recognizing and responding to this global dimension of the risk calls for interdisciplinary and international dialogue to harmonize guidelines, research, surveillance methods and the response in case of adverse outcomes. National registries will need to be compatible, and archived tissue perhaps accessible to scientists from other countries. Many of these issues were agreed upon at the World Health Organization consultation on xenotransplantation in Geneva last October.

With traditional research funding drying up in universities, biotechnology companies are increasingly involved in cutting-edge research. Transparency through peer-reviewed publication cannot be guaranteed and, although companies linked to major pharmaceutical companies will probably be guided by long-term interests, some venture-capital companies may seek short-term rewards. The long-term effects are unpredictable but, given that the public is being put at risk, an argument can be made for greater public debate, as suggested by Bach *et al.*², and perhaps novel means of public supervision.

A challenge in the days ahead is to define these means and to define what might constitute public or community consent.

Xenotransplants will almost certainly be too expensive in the early years to be the answer to the shortage of human cadaveric organs for transplantation. Furthermore, it is not yet known if xenogeneic organs will adequately replace the function of complex metabolic organs such as the liver. We must therefore continue to strive for other means of increasing donation of human organs and certainly ensure that xenotransplantation does not undermine allo-donation.

Xenotransplantation may also affect traditional transplant ethics³ by eroding the 'gift' metaphor and the confidentiality principle; through the complete reification and commodification of organs and the rewriting of the meaning of consent to include the community; and through erosion of the courtesy and cooperation between centres working with a very scarce resource.

Finally, no sensible person would want to hold back development of such a potentially useful technology were it not for the risk to public health. There may be reason to be confident that the risk of xenozoonoses is small, but the risk would be justified only if large numbers of patients could be saved in the very near future and we had no hope of improving our risk assessment capabilities quickly. This is not the case at present: xenotransplantation will have no immediate effect on overall transplant numbers; and risk assessment is improving.

We understand the similarity to the early days of genetic engineering and the call for an embargo then, which has proved unnecessary. However, one similarity should not constitute a cliché; the risk is serious enough and the immediate benefits are small enough to require this situation to be assessed in its own right. This is particularly true because, although most comments have been about clinical trials, the reality is that even minimal perception of success in trials will soon lead to xenotransplants as therapy and will encourage many unprepared units to 'do' xenotransplants. The immediate effect of the permissive US PHS guidelines will be that other countries will want to follow suit, so speeding further a process that the PHS itself felt obliged only a few weeks ago to halt as a result of new evidence on porcine endogenous retroviruses.

The PHS is listening and responding. Its guidelines are, fortunately, still evolving. Common sense would dictate caution even without a formal embargo. The guidelines

are tougher than originally envisaged and may become tougher still. It would be a recognition of its global reach for the US PHS to acknowledge that it is here dealing with a global rather than a US issue and that there is no reason to hurry past prudence.

A. S. Daar

Sultan Qaboos University,
Muscat, Oman
e-mail: asdoc@gto.net.om

Sir — We, the American Society of Transplant Physicians (ASTP) and the American Society of Transplant Surgeons (ASTS), reject the call for a moratorium on clinical trials of xenotransplantation in the United States². Bach *et al.* assert that the potential infectious risks of xenotransplantation could create a public health problem, that a broad public discussion is therefore required and that a moratorium is justified until the latter is completed. But their opinions represent, at best, a minority among US transplant professionals.

They also appear to ignore the remarkable public discussion organized over the past four years by US Public Health Service (PHS) agencies, including the Food and Drug Administration (FDA), the Centers for Disease Control and the National Institutes of Health. Public meetings, the most recent in January, have brought together hundreds of professionals in transplantation medicine and surgery, infectious disease, veterinary medicine, ethics, public policy and law, as well as patients, families, animal rights representatives and companies. This knowledge base has been shared at meetings organized by the Institute of Medicine (Washington, June 1995), Health Canada (Ottawa, November 1997) and the World Health Organization (Geneva, October 1997). All the issues raised by Bach *et al.* have already been clearly articulated, widely discussed and published^{1,4–6}, and have received substantial media coverage. Between November 1996 and October 1997, some 230 articles discussing xenotransplantation were published in major newspapers in the United States, while more than 435 were published in the UK national press in 1996–97.

The consensus that has emerged from the PHS discussions is that it is time to proceed cautiously with well-defined and highly controlled clinical trials. To support this process further, the PHS has developed several important mechanisms, including a Xenotransplantation Advisory Subcommittee and plans for a National Patient Registry, Biological Specimen

Repository and a National Advisory Committee. We strongly endorse this consensus opinion, and commend the PHS for its efforts in support of our patients, their families and the general public. We agree fully with the recent public statements of William Raub (Deputy Assistant Secretary of Health and Human Services) and Michael Friedman (Acting Commissioner, FDA) rejecting the call for a moratorium.

Daniel R. Salomon

(Xenotransplantation Committee, ASTP)

Department of Molecular
and Experimental Medicine — SBR5,
Scripps Research Institute,
La Jolla, California 92037, USA
e-mail: dsalomon@scripps.edu

Ronald M. Ferguson

(President, ASTS)

Ohio State University Medical Center,
Columbus, Ohio

J. Harold Helderman

(President, ASTP)

Vanderbilt University Medical Center,
Nashville, Tennessee

Sir — As a clinician involved in clinical allotransplantation and experimental xenotransplantation (heart-lung and lung), I should like to comment on your attitude to xenotransplantation¹.

The European Union has provided a

ECU1.5 million (US\$1.6 million) grant (1997–2000) to establish a working group of several non-UK countries to address different topics of xenotransplantation. The unique aspect of this group is that it is composed of a physician, researchers (in biochemistry, immunology, genetics and virology), a sociologist, philosophers and so on.

Although heterogeneous, this group has a unique question in mind: is xenotransplantation of clinical benefit for a human being? Although you have published quite a lot on this subject, as far as I know none of the authors has been at the bedside of a patient. They cannot feel the frustration of patients who die while waiting for an organ nor the wonderful sensation of being able once more to breathe or to move without effort. What about the parents seeking a therapeutic solution for children with terminal diseases? As a clinician, is it really ethical for me to have no solution or should I be more concerned with infection and social arguments?

In the laboratory in which I work, we have recently used the nude mouse to host a human trachea derived from human embryonic cells. After several months, we transplanted up to 10 cm³ of human trachea into piglets to test whether the process might be used in human babies as an

alternative to their death. All the experiments were successful (without immunosuppression, human grafts were not rejected). Do I have the right to propose this technique to the parents? I think I do and we are in the process of asking for permission to do it.

I believe that people dealing with xenotransplantation should give more thought to the patient rather than to ways of raising funds for research that may never be applied in clinical practice because it is too complicated. Have such researchers any idea what a child looks like after several years of immunosuppression treatment?

The time has come for clinicians rather than basic researchers to give their opinions on clinical xenotransplantation.

Paolo Macchiaroni

Department of Thoracic
& Vascular Surgery,
Hôpital Marie-Lannelongue,
Paris-Sud University,
133 Avenue de la Résistance,
F-92350 Le Plessis Robinson, France
e-mail: paolo@pratique.fr

1. Butler, D. *et al.* *Nature* **391**, 320–325 (1998).
2. Bach, F. H. *et al.* *Nature Med.* **4**, 141–144 (1998) & *Nature* **391**, 326 (1998).
3. Daar, A. S. *World J. Surgery* **21** (9), 975–982 (1997).
4. Chapman, L. E. *et al.* *New Engl. J. Med.* **333**, 1498–1501 (1995).
5. Institute of Medicine, *Xenotransplantation: science, ethics and public policy* (National Academy Press, Washington, DC, 1996).
6. Weiss, R. A. *Nature* **391**, 327–328 (1998).

Search for consensus on genetic engineering

Sir — If the Genschutz-Initiative (gene-protection initiative) is accepted by a majority of Swiss voters in June (see *Nature* **391**, 312; 1998), a constitutional change will lead to the strictest regulation of gene technology anywhere in the world. It may appear astonishing that the Swiss — whose wealth depends so heavily on research and development leading to new technologies — would accept such restrictions on the use of what is likely to become the key technology of the future.

Perhaps a more prudent implementation of gene technology, and a better prepared social and ethical discussion of its desirability, could have prevented the successful signature campaign for such a restrictive initiative. (In Switzerland, at least 100,000 voters have to support an initiative to have a referendum on a constitutional change.) But the pace at which applications of gene technology are implemented makes either of these difficult to achieve. Indeed, it is questionable whether existing political instruments are adequate to conduct an effective discussion of such new technologies.

The only reasonable alternative is to intensify the social and political discussion. For this purpose, new institutions are needed, and experts trained to communicate information to lay people about new technologies and their potential implications in a neutral environment, without hierarchical structures and patronizing attitudes.

It also requires flexible political instruments, such as 'consensus conferences', in which groups whose members are selected as representative of age, gender, education and occupational skills cross-examine expert witnesses and discuss such topics. Although their conclusions — often surprisingly sophisticated, despite a short preparation time — are not legally binding, they can serve as a useful political barometer to politicians and institutional decision-makers.

'Consensus conferences' are not a solution to all political problems. But they could be an important instrument in the future, particularly in Switzerland, where there has so far been reluctance to use such political instruments.

Michael Altmann

*Institut für Biochemie
und Molekularbiologie,
Universität Bern, Bülhstr. 28,
3012 Bern, Switzerland
e-mail: michael.altmann@mci.unibe.ch*

Growing benefits of biotechnology

Sir — In his review of my book *The Doubly Green Revolution*¹, N. W. Simmonds asserts that biotechnology's "plant breeding achievements so far are nil". This is simply not true. Even the average supermarket shopper is aware that the products of genetically engineered crops are on the shelves. Last year, more than 20 million acres of such crops were harvested in the United States, and there have been some significant achievements in developing countries².

Tissue-culture techniques in China and South Korea have produced several new rice varieties; one, La Fen Rockefeller, is giving up to 25% higher yields. Transgenic rices are now available incorporating resistance to insects, bacterial blight, rice stripe virus and hoja blanca virus. Molecular markers have been used to incorporate multiple resistance genes in rice in China, India and Colombia. None of this would have been possible without biotechnology.

Gordon Conway

*Sussex House, University of Sussex,
Falmer, Brighton BN1 9RH, UK*

1. Simmonds, N.W. *Nature* **391**, 139 (1998).

2. James, C. *ISAAA Briefs* No. 5 (Intl Service for the Acquisition of Agri-Biotech Applications, Ithaca, New York, 1997).

Benefits and dangers of genetic tests

Sir — The potential for causing mischief by carrying out genetic analysis on an estimated 283 million archived tissue samples has become the focus of the US National Bioethics Advisory Commission.

French-Canadians living in Maine or Japanese-Americans living in California might learn about increased susceptibility to certain genetic traits, including disease predispositions, without any member from these groups consenting to participate in such analyses. Similarly, tribal customs might be disregarded in handling surgical or post-mortem remains. This had led to proposals for "community consultation"¹.

In a News story, "'Group debate' urged for gene studies"², Meredith Wadman cited a study of increased heritable susceptibility to colon cancer among Ashkenazi Jews as the trigger for this recommendation³. I should like to emphasize that, rather than ignoring the Ashkenazi Jewish community, this study has been a model for community consultation.

The study was carried out following population genetic studies of heritable susceptibility to Canavan's disease and to breast cancer (from inheritance of the *BRCA2* 6174delT mutation) among Ashkenazi Jews^{4,5}. The participants in these studies were Ashkenazi Jewish people who had sought carrier testing for Tay-Sachs disease and cystic fibrosis. Approval of the institutional review board was obtained and subjects provided specific consent for anonymous population genetic studies of heritable susceptibility to these conditions. Samples were rendered anonymous upon receipt in the laboratory so that they could no longer be linked to their identifiers.

The results of the *BRCA2* study received widespread coverage in the press and on television. Results of this study were presented at national scientific meetings and at leadership meetings of Jewish organizations. Letters about heritable susceptibility to disease among Jews were sent to synagogues and Jewish community centres. In the face of all this publicity, individuals have continued to provide their consent for such analyses both before and after the colon cancer study, and 1,600 DNA samples have been accumulated. Indeed, as the director of the genetic screening programme and the population genetic registry upon which these studies were based, I can assume only that the participants who spoke to physicians and counsellors in my programme and provided informed consent were acting as representatives of their community.

The intention of these studies was to

benefit rather than create mischief for the Ashkenazi Jewish community. Because the colon cancer study was based on a relatively small sample and because of concern about unfair genetic discrimination, I expressed caution about lowering the threshold for offering genetic testing until the risks of being a carrier were more precisely quantified⁶. On the other hand, carrier screening for Canavan's disease has become standard. The usefulness and limitations of heterozygote testing for *BRCA2*, along with *BRCA1*, have been validated in a series of other studies for people with a family history of breast cancer.

Throughout this time, serving the needs and maintaining the trust of the Ashkenazi Jewish community have been the paramount concerns of my programme.

Harry Ostrer

*Human Genetics Program,
NYU Medical Center,
550 First Avenue, MSB 136,
New York, New York 10016, USA
e-mail: ostrer01@mcrcr6.med.nyu.edu*

Sir — I read with interest and dismay the News article about the recommendations on informed consent and the ethical use of tissue samples to be issued by the US National Bioethics Advisory Commission².

There is no doubt that the use of archival tissues for genetic epidemiological and pathological studies carries with it complicated ethical issues — for example, who should profit from the discoveries made using this material? One of the arguments put forward to restrict scientific access to these samples, however, is that the information creates job and insurance risks for the individuals or populations analysed. Clearly, job and insurance discrimination are potential dangers associated with this type of research and there are many reasons to argue that these should be avoided.

However, the approach of restricting scientific access to archived tissue samples is not the way to do it. In many ways this is like saying that we should prevent stores from selling sharp knives because they might be used to harm someone. Knives can be very useful or very dangerous, depending on the use. The recommendations and legislation should be aimed not at the scientists, but at the uses that we consider unethical. Many of us believe that insurance discrimination is one such area. Perhaps our efforts would be better spent considering the ethics of the insurance practices that will be affected by the new information.

Most scientists are using these archived samples to investigate disease pathogenesis. This knowledge can be used to develop better diagnostic and treatment strategies to help the individuals and populations afflicted by the disease. Indeed, the

populations affected have the most to gain from the scientific use of their specimens.

The bottom line is that the scientific use of archived samples is good for all of society. This fact will not change no matter how many different communities are consulted. If this is considered unethical, maybe we should re-examine our ethics rather than hinder the progress of biomedical research.

Richard H. Scheuermann

Louis J. Picker

*Department of Pathology,
University of Texas, Southwestern Medical Center,
Dallas, Texas 75235-9072, USA
e-mail: scheuermann.richard@
pathology.swmed.edu*

1. Foster, M. W., Eisenbraun, A. J. & Carter, T. H. *Nature Genet.* **17**, 277–279 (1997).
2. Wadman, M. *Nature* **391**, 314 (1998).
3. Laken, S. J. et al. *Nature Genet.* **17**, 79–83 (1997).
4. Am. J. Hum. Genet. **57**, 1250–1252 (1995).
5. Oddoux, C. et al. *Nature Genet.* **14**, 188–190 (1996).
6. *New York Times* 26 August 1997.

About face

Sir — Although I am grateful that my book *About Face* was reviewed in *Nature* (**390**, 458; 1997), I should like to point out that, contrary to Stuart Sutherland's claim, I was not instrumental in setting up the charity Changing Faces.

Sutherland also says that I suggest that the emotional problems in autism stem from an inability to interpret others' facial expressions. I find it difficult to see how he arrives at this simplistic conclusion. In the book I write: "But autism, of course, is a widespread developmental disorder, not just a face thing. Do the many other problems [autistic people] have preclude too many conclusions about relatedness and social development based on facial action? Perhaps. But what if there were people with facial problems alone, and what if they experienced some similar difficulties in relatedness? Autism is a condition with problems in social interaction which are reflected in their problems with the face. What if we could turn this upside down? How might facial problems affect social development and selfhood?"

Although coming at the end of the chapter on autism, these questions are used as a bridge to refer to the next chapter on Möbius syndrome (in which people have congenital lack of facial movement). Sutherland seems to have missed this link.

Indeed, I write later: "Autism of course is not a face problem in the way Möbius Syndrome is".

Jonathan Cole

*Hollyhock Cottage, Minstead,
Lyndhurst,
Hampshire SO43 7GG, UK
e-mail: jcole@minstead.demon.co.uk*

A paradigm gets shifty

W. Ford Doolittle

The prevailing view of the origin of complex eukaryotic cells presumes a symbiosis, based on respiration, between a bacterium and a primitive eukaryotic 'host'. But could the symbiosis have been based instead on hydrogen metabolism, with the host being an archaean?

If we look at cell structure, there are clearly only two kinds of organisms in the living world: prokaryotes (Bacteria and Archaea) and eukaryotes (everything else). The differences that mark this dichotomy are shown in Fig. 1. In the main, eukaryotic cells have complex structural features (such as the membrane-enveloped true nucleus, or eu-karyon, which gives the group its name) and prokaryotic cells don't. Many eukaryotic and some prokaryotic organisms are multicellular, but there are no difficulties in distinguishing them at the level of cell structure, and no obvious transitional forms.

Not surprisingly, biologists have long assumed that eukaryotes evolved from prokaryotes, and have come up with theories for the origin of the more complex eukaryotic structural components that involve particular selective advantages, mechanisms and histories. The best articulated and most widely accepted of the theories, 30 years old this year, is Lynn Margulis's version of the 'endosymbiont hypothesis' — 'endo-' for internal, and 'symbiont' as in a partner in a mutually beneficial relationship — for the origin of mitochondria and plastids¹. Mitochondria are the powerhouses of most eukaryotic cells, generating ATP in respiration; plastids (chloroplasts) are the organelles of photosynthesis in algae and plants.

In textbook accounts, the endosymbiont theory takes in several notions, as follows.

- The eukaryotic endomembrane (including the nuclear membrane) and cytoskeletal systems arose in some ancestral anaerobic prokaryotic lineage, and were selected for because these systems made the engulfing of solid particles possible.
- Cells of the resulting 'proto-eukaryotic' lineage made a living by engulfing and digesting prokaryotes.
- Sometimes these prokaryotes, rather than being digested, became endosymbionts, conferring a metabolic advantage on their predator, which then became their host.
- An endosymbiosis between an aerobic bacterium (on biochemical grounds, probably an α -proteobacterium similar to the present-day *Paracoccus*) and an anaerobic proto-eukaryotic host was stabilized

through mutual advantage — respiration-derived ATP in exchange for metabolizable substrates and physical protection.

- Genes needed only for independent growth were lost from the genome of the endosymbiont, while some endosymbiont genes needed directly or indirectly for respiration were transferred to the host nucleus.

- The result for the endosymbiont was to become a modern mitochondrion, an organelle with a highly reduced genome (coding for only about a dozen proteins in animals). Similarly, but not discussed further here, chloroplasts evolved from endosymbiotic cyanobacteria.

Most people now consider that molecular sequence data, accumulated between about 1975 and 1995 (ref. 2), clinched the endosymbiont hypothesis for mitochondria. Such data showed that genes in mitochondrial genomes (together with genes in eukaryotic nuclear genomes whose products

are targeted to mitochondria) are indeed of bacterial, and gratifyingly of α -proteobacterial, origin, while indicating that the rest of the eukaryotic nuclear genome may be of archaea-like ancestry³. Excitingly, components of the machinery for the replication, transcription and translation of DNA that were once thought to be eukaryote-specific, are now known from archaeans⁴. So we have come to believe in the endosymbiont hypothesis as outlined in Fig. 2 (overleaf; see also ref. 5).

However, the sequence data only address the origins of genes: they have been taken as support for the assumptions about mechanism and selective advantage outlined above primarily because there was no other well-fleshed-out theory which made these same predictions about genes. Data published over the past two or three years, much of them from genome-sequencing projects, have hinted that it is time for a new theory. In particular, it is turning out that eukaryotic nuclear genomes carry many genes of bacterial (sometimes α -proteobacterial) origin which have nothing to do with mitochondrial functions. Moreover, mitochondrion-free eukaryotes that we had come to think of as direct descendants of ancient proto-eukaryotes carry mitochondrial genes in their nuclear genomes.

It is possible to cobble together an expanded version of Margulis's theory to accommodate all of this⁶. But, on page 37 of this issue⁷, William Martin and Miklós Müller take a much bolder step. They pro-

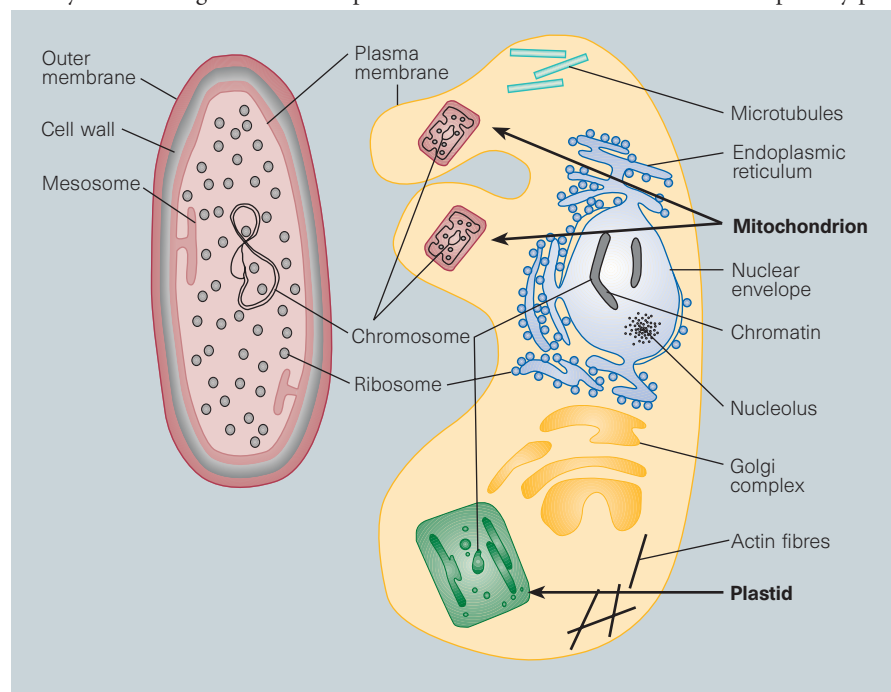


Figure 1 Typical prokaryotic (left) and eukaryotic (right) cells. In prokaryotes, the single, circular chromosome is in contact with the cellular cytoplasm (cytosol). In eukaryotes, highly structured, linear chromosomes are (much of the time) separated from the cytosol by a nuclear envelope, which is itself part of a complex internal membrane system unique to eukaryotes. The eukaryotic cytoskeleton, of which actin fibres and microtubules are part, provides an internal prop and mediates intracellular movement.

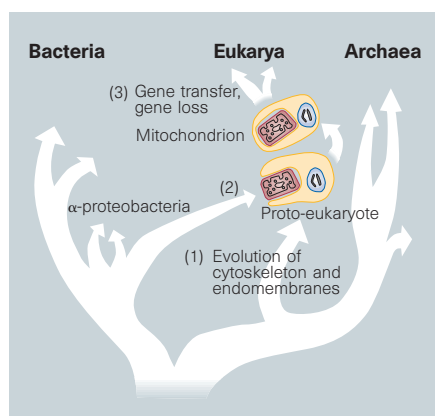


Figure 2 The endosymbiont hypothesis for the origin of mitochondria, as currently construed. The deepest branching in the Tree of Life separates bacteria from archaeans and eukaryotes. The latter evolved from an archaean-like ancestor. Three crucial steps are shown: (1) the evolution of cytoskeletal and endomembrane systems, driven by selection for the ability to engulf solid food particles such as bacteria; (2) the conversion of undigested α -proteobacteria to intracellular respiratory symbionts ('endosymbionts') and then into mitochondria; and (3) the loss of many of the original α -proteobacterial genes and the transfer of others to the nucleus.

pose a model (the 'hydrogen hypothesis') which is in some ways more plausible than the current view. It is testable in several of its implications; and, whether true or false, it is the first new hypothesis about eukaryotic origins in 30 years to have been really thoroughly articulated at the biochemical, molecular and cellular levels.

What Martin and Müller propose is this.

- There was indeed an ancient symbiosis involving an α -proteobacterium as one partner and a relative or member of the Archaea as the other. The relevant activity of the α -proteobacterium was not, however, respiration, as in Margulis's hypothesis, but the excretion of H_2 and CO_2 — which are the waste products of anaerobic fermentation of externally available, reduced organic compounds in some contemporary α -proteobacteria. The archaean partner, like modern methane-producing archaeans, used H_2 and CO_2 as its sole sources of energy and carbon. Both metabolisms could go on simultaneously and independently in anaerobic environments like many known today, and there are many examples of mutual relationships based on their coupling.

- In the absence of an outside source of hydrogen, the archaean (nominally the host) became dependent on the α -proteobacterium (the symbiont). Selection on host genes produced tighter and tighter physical association and greater surface contact between the two.

- Gene transfer from symbiont to host pro-

vided the latter with membrane proteins for the import of the substrates and enzymes for glycolysis, the process by which ATP is generated anaerobically. So the host could begin to feed its symbionts, and thus surround them completely.

- The effect of this was conversion of the host from autotrophy (using H_2 and CO_2 as substrates) to heterotrophy (able to use complex organic molecules). The symbiont was then either lost, converted to a hydrogenosome, or became a mitochondrion (in the evolutionary line leading to all complex cells).

The reader may well complain that, mitochondria apart, Martin and Müller have not addressed the origin of the other complex eukaryote-specific features shown in Fig. 1. But neither do textbook versions of Margulis's endosymbiont hypothesis — the strongest support for which remains the presence of α -proteobacterial genes in eukaryotic genomes. The hydrogen hypothesis accounts for this α -proteobacterial presence, and for the kinds of genes concerned, more naturally than does the endosymbiont theory. It more readily explains the energy metabolism of eukaryotes that lack mitochondria (which, however, were little known when Margulis formulated her hypothesis). And by viewing bacterium-to-archaean gene

transfer in terms of ecological interactions, it also hints at an explanation for the many bacterial genes in the genome sequence of the heterotrophic archaean *Archaeoglobus*⁸, published late last year.

None of this is to say that Martin and Müller are right: the classical endosymbiosis theory may yet be amended and saved. But the onus is on its defenders to say why the hydrogen hypothesis is not at least as good. In the genomes that have been or will shortly be completely sequenced, we may already have all the data we need. We need the wit — and the will — to analyse them open-mindedly. □

W. Ford Doolittle is with the Program in Evolutionary Biology, Canadian Institute for Advanced Research, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada.
e-mail: ford@is.dal.ca

1. Margulis, L. *Origin of Eukaryotic Cells* (Yale Univ. Press, New Haven, CT, 1970).
2. Gray, M. W. & Spencer, D. F. *Symp. Soc. Gen. Microbiol.* **54**, 109–126 (1996).
3. Doolittle, W. F. & Brown, J. R. *Proc. Natl Acad. Sci. USA* **91**, 6721–6728 (1994).
4. Olsen, G. J. & Woese, C. R. *Cell* **89**, 991–994 (1997).
5. Margulis, L. *Proc. Natl Acad. Sci. USA* **93**, 1071–1076 (1996).
6. Doolittle, W. F. *Symp. Soc. Gen. Microbiol.* **54**, 1–21 (1996).
7. Martin, W. F. & Müller, M. *Nature* **392**, 37–41 (1998).
8. Klenk, H.-P. et al. *Nature* **390**, 364–370 (1997).

Solar System

Two shades beyond Neptune

Clark R. Chapman

We know that the composition of bodies in the Solar System changes with distance from the Sun. But it is unexpected that, as reported on page 49 of this issue¹, there may be two basic compositions of bodies at the outer extremity of the planetary system.

An obvious compositional divide separates the terrestrial planets, made of rocks and metals, from the outer, giant planets, made of ices and gases. More subtle differences separate asteroids, which range from pure metal (such as 16 Psyche) to fluffy, carbonaceous-meteorite-like material (such as 253 Mathilde, examined last June by the Near Earth Asteroid Rendezvous spacecraft²). This variation between asteroids was revealed in the 1960s by measuring the colour of reflected sunlight: the metals and silicates of most inner-belt asteroids make them slightly reddish, whereas the carbonaceous outer-belt asteroids are neutral in colour, and very dark.

The asteroid belt is evidently a transition region between the hot, inner Solar System, where high-temperature condensates accreted into small bodies and eventually into rocky planets, and the colder, outer Solar System where volatile ices condensed, along with high- and low-temperature minerals.

Asteroid colours may also reflect subsequent evolution, such as the early segregation of minerals within some bodies that were heated and melted, and surface weathering by solar-wind particles and meteoroids.

Small bodies in the outer Solar System have been mostly out of reach for analogous studies. (Although comets observed in the inner Solar System come from outer-Solar-System reservoirs, their comas hide the properties of their nuclei.) These bodies are far away from both us and the Sun, so they are extremely faint, even in the largest Earth-based telescopes. Indeed, it was only six years ago that David Jewitt and Jane Luu discovered³ the first trans-neptunian object besides Pluto and Charon, called 1992 QB1.

Since then, about 60 more so-called Kuiper-belt objects (KBOs) have been found⁴. These primordial-remnant bodies, up to a few hundred kilometres in diameter, form a torus that lies between 35 and at least 50 AU from the Sun (1 AU is the mean Earth–Sun distance). They confirm the hypothesis of K. E. Edgeworth and G. P. Kuiper half a century ago that the original disk-like solar nebula of gas and dust did not stop abruptly at Neptune⁵. Small bodies presumably formed everywhere in the early Solar

System, but in general they were too close to a planet for their orbits to remain stable for billions of years⁶ — except in the gap between Mars and Jupiter, and beyond Neptune.

Seven bodies, termed Centaurs, have also been found orbiting chaotically in the outer Solar System. They are probably recent escapees from the Kuiper belt. According to numerical simulations, they will gradually move inwards to become short-period comets^{7,8}.

Spectrophotometry has improved enough in the past few years for us now to be able to measure the colours of some KBOs, and of the known Centaurs. Three teams of observers — Jewitt and Luu; a group led by S. F. Green; and S. C. Tegler, W. Romanishin and colleagues — are responsible for most of the published colours. KBO colours vary quite widely.

I should note that the colours of KBOs and asteroids are subtle. Although colour differences are real, reflecting profound mineralogical differences, all asteroids and KBOs would look sensibly grey to the eye. The Moon, surprisingly, is quite reddish by astronomical standards. Even the Centaur 5145 Pholus, often called “extraordinarily red” by astronomers, would look greyish-brown at best and is far less coloured than, say, Mars.

The new surprise is that the largest sample of KBOs and Centaurs studied to date, 16 in all, divide cleanly into two colour groups. Using the Steward Observatory 2.3-m telescope in Arizona, Tegler and Romanishin¹ find one group with neutral colours, like the carbonaceous asteroids, and another group with much redder colours. There is nothing in between.

Jewitt and Luu⁹ have tentatively suggested that KBO colours may vary systematically with body size, based on their independent data for five KBOs. (Their less thorough observations are systematically offset from the colours of Tegler and Romanishin, which indicates the difficulty of observing objects so faint, but are consistent with the two colour groups.) But Tegler and Romanishin find, in their larger sample, no correlation between colour and any physical or orbital property.

What might explain the two groups? There are processes, such as long-term irradiation interrupted by collisions that create fresh surfaces, that can change colours — but how could they explain the absence of bodies with intermediate colours? The KBO bimodality reminds us of the asteroid bimodality, which reflects the inner-to-outer-planet transition zone. Yet no analogous transition is thought to exist in the Kuiper belt and, anyway, there is no correlation among KBOs between colour and distance from the Sun.

The asteroidal bimodality was sharpened by the fact that some asteroids went through

a process of heating and melting, whereas others didn't, for reasons that remain puzzling. Could the same thing have happened to the KBOs? One source of asteroid heating, the decay of short-lived radionuclides like aluminium-26, would probably not have affected Kuiper-belt objects, which accreted more slowly. For now, all we can say is that the two distinct colour groups may betray something important and unexpected that happened long ago in the deep-freeze of the outer Solar System. □

Palaeobiology

Mass extinction probed

Charles R. Marshall

At the end of the Cretaceous period, 65 million years ago, a large meteorite or comet struck the Earth. Coupled, perhaps, with longer-term climatic changes, this led to a mass extinction — with the disappearance of 39–47% of fossilizable genera and perhaps 75% of species. What were the secondary causes that translated the bolide impact into a mass extinction? On page 69 of this issue¹, Smith and Jeffery report their detailed study of the sea-urchin fossil record across the Cretaceous–Tertiary (K/T) boundary, and provide insights into the nature of these causes.

The fossil record reveals two modes of species extinction — low levels of relatively continuous background extinction, punctuated by occasional large mass extinctions (Fig. 1a, page 19). Mass extinctions account for only a few per cent of all species extinctions². However, they are disproportionately significant because they removed evolutionarily resilient ‘incumbent’ groups, such as the dinosaurs, opening brief windows of opportunity for other groups to become dominant (the mammals after the demise of the dinosaurs, for example). Understanding the causes of these mass extinctions is critical to explaining the history of life.

Smith and Jeffery probed the secondary causes of the end-Cretaceous mass extinction by searching a global database for biological and geographical correlates of sea-urchin survival and extinction across the K/T boundary. Distilling the required data from over 150 years of published studies is not straightforward. For example, fossils that belong to one species may have been assigned to several different species, especially by palaeontologists working in different geographical areas and in rocks of different ages. This will artificially inflate perceived levels of extinction. Furthermore, the intensity of the extinction may be overestimated if different species on either side of the extinction boundary are not recognized as belonging to the same evolutionary lineage. How best to correct for this type of error is not settled³.

Clark R. Chapman is at the Southwest Research Institute, Boulder, Colorado 80302, USA.
e-mail: cchapman@boulder.swri.edu

1. Tegler, S. C. & Romanishin, W. *Nature* **392**, 49–51 (1998).
2. Veverka, J. et al. *Science* **278**, 2109–2114 (1997).
3. Jewitt, D. C. & Luu, J. X. *Nature* **362**, 730–732 (1993).
4. <http://cfa-www.harvard.edu/iau/Ephemerides/Distant/index.html>
5. Stern, S. A. *Astron. Soc. Pacif. Conf. Ser.* **107** (eds Rettig, T. W. & Hahn, J. M.) 209–232 (Astron. Soc. Pacif., San Francisco, 1996).
6. Gladman, B. & Duncan, M. J. *Astron. J.* **100**, 1680–1693 (1990).
7. Stern, A. & Campins, H. *Nature* **382**, 507–510 (1996).
8. Levison, H. F. & Duncan, M. J. *Icarus* **127**, 13–32 (1997).
9. Jewitt, D. & Luu, J. *Astron. J.* (in the press).

To combat these problems, Smith and Jeffery used data from the literature only where they could confirm the identity and morphology of species, and they carried out phylogenetic analyses to determine which species belonged to the same evolutionary lineage. Using these quality controls, they found that the extinction intensity of sea-urchin genera across the K/T boundary was about 36% — considerably less than a previous uncorrected estimate of around 70% (ref. 4). But even this new estimate may be too high. First, some of the genera that were last seen in the final stage of the Cretaceous (the Maastrichtian) probably became extinct before the impact at the end of this period. We do not yet have enough temporal control or detailed stratigraphic data to dissect out the actual times at which most genera within this stage became extinct. Certainly for other groups, however, some of the species that became extinct in the Late Maastrichtian disappeared a considerable time before the K/T boundary⁵. Second, there is a dearth of rocks bearing fossil urchins just after the K/T boundary⁶. More survivors of the Cretaceous may be found as more post-Cretaceous rocks are sampled, although the lack of rocks may represent a true loss of suitable habitat and diversity.

To what extent can species be treated as equivalent entities in the face of mass extinction? Can they be treated as interchangeable ‘particles’? Or did their unique characteristics dictate their survival, and, if so, at what level of generality? If species can be treated as particles, extinction intensity might be expected to correlate with the number of species in each genus, or with their geographical ranges. But Smith and Jeffery found that sea-urchin extinction intensity is not correlated with these characteristics. Other groups do not usually show a connection with species richness, but the lack of correlation with geographical range is unexpected^{7–9}. The urchins do show a hint of a link with water depth, suggesting that we may find that urchins in different habitats

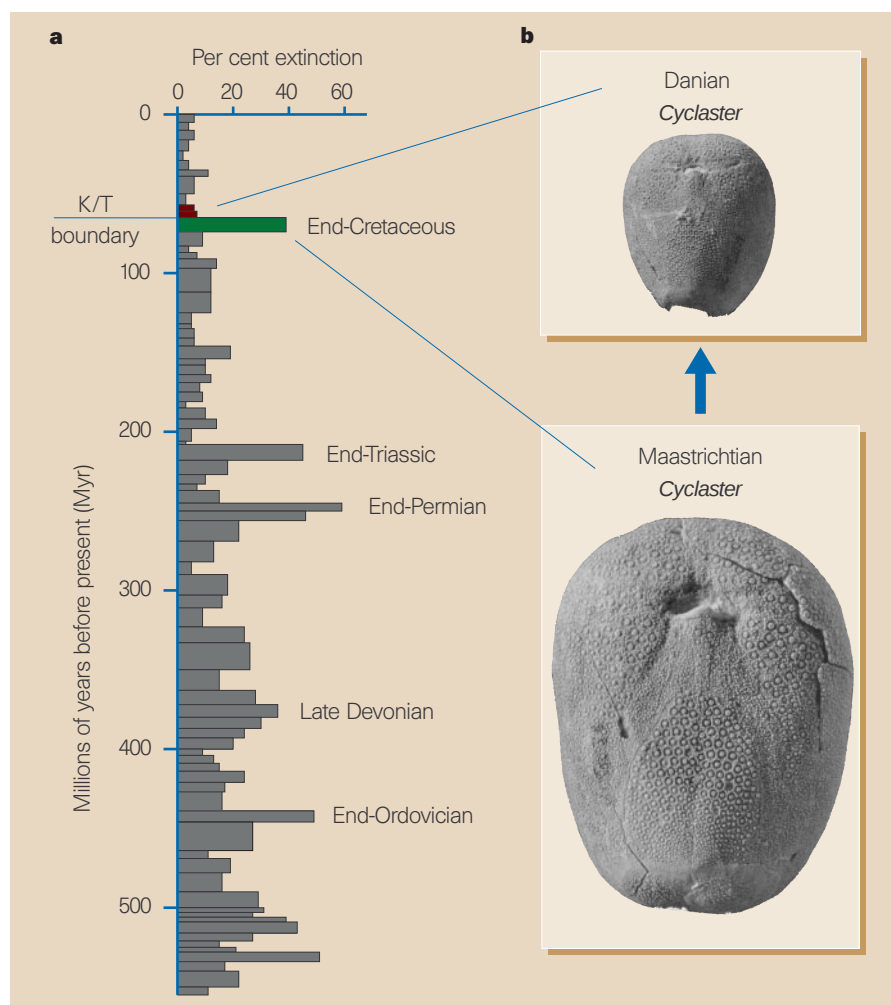


Figure 1 The mass extinction at the Cretaceous–Tertiary (K/T) boundary. **a**, The extinction intensity of all marine animal genera measured for each stage of the Phanerozoic¹³. **b**, Smith and Jeffery¹ have found that sea-urchin lineages that survived the end-Cretaceous mass extinction, one of the ‘Big Five’ mass extinctions, showed a dramatic drop in body size. The sea urchin *Cyclaster*¹⁴ is shown, the Danian specimen being 2.1 cm in length and the Maastrichtian 3.8 cm.

were more severely affected (or not) than others, independent of the biological characteristics of the urchins in each habitat. As with other groups^{7,8}, the authors found no strong correlation with broad biological characteristics, such as whether or not the species lived within the sediment.

Tantalizingly, Smith and Jeffery found that the survival of genera restricted to one region depended strongly on the region. The Americas — posited to have been hardest hit by the immediate impact blast¹⁰ — had fewest survivors, whereas more distant regions were less disturbed. Perhaps the authors have discovered a signature of the impact blast itself, which occurred at Chicxulub in Mexico. However, the pattern may instead reflect the lack of rock types appropriate for fossil preservation in key geographical areas.

The strongest correlate of survival was with the type of sea urchin involved — differences in their biological characteristics played a big part in the fate of the urchins. Specifically, there was a correlation with

feeding mode, suggesting that this (or other biological characteristics that correlate with feeding mode) was involved in their survival. Five main feeding types are discernible in fossil urchins. Those that lived in nutrient-poor settings and fed on organo-detritus, without the aid of specialized feeding structures such as tube feet, fared the worst. Generalist omnivores suffered the least extinction.

The most easily interpreted signal in the sea-urchin data is seen when extinctions are ignored. The authors found a pronounced drop in adult body size for most lineages that survived the K/T boundary (Fig. 1b). These data indicate that there was a sustained drop in primary productivity after the K/T boundary, a conclusion also reached by others^{11,12}. Combining this conclusion with their other data, Smith and Jeffery suggest that sea-urchin extinction was driven mainly by a nutrient crisis, and that the adult life stage was more severely affected than the larval stages. But the severity of the extinction seems to have depended on the way in which



100 YEARS AGO

In collecting the literature regarding phosphorescent plants, I chanced on an article, by Mr. C. F. Holder, on “Living Lamps,” in No. 392, vol. lxvi. (January 1883) of Harper’s Magazine. In this article, at page 191, it is said: “In South America, a vine known as the Cipo, when injured, seems to bleed streams of living fire. Large animals have been noticed standing among its crushed and broken tendrils, dripping with the gleaming fluid, and surrounded by a seeming network of fire.” Could any reader of *NATURE* confirm the existence of this Cipo with a phosphorescent sap? Cipo, I believe, is a name for liana, not for vines. If true, the existence of a phosphorescent sap in a superior plant would be of great physiological interest. But no mention of this or a similar case is to be found in the standard works on vegetable physiology. I fear the statement may have as much foundation as the assertion, made in the same article, that among the peasantry of Italy girls complete their gala toilet with diadems of fireflies. — Italo Giglioli, Portici, near Naples, February 18. From *Nature* 3 March 1898.

50 YEARS AGO

An interesting appointment has been made at Leicester to the newly created chair of mathematics. Dr. R. L. Goodstein, of St. Paul’s School and Magdalene College, Cambridge, has been a lecturer at Reading since 1935. He has no less extensive a knowledge of mathematics, as commonly understood, than any of his contemporaries; but it was Wittgenstein who excited and inspired him, and he has made his reputation and won his position by original work on the foundations of mathematics. It is half a century since Russell first insisted that investigation into the principles of mathematics was a task for expert mathematicians, not for philosophers, but hitherto the subject has been one in which a mathematician could not take more than an amateur interest without endangering his professional status. A bad tradition has been shattered at last.

From *Nature* 6 March 1948.

Many more abstracts like these can be found in *A Bedside Nature: Genius and Eccentricity in Science, 1869–1953*, a 266-page book edited by Walter Gratzer. Contact David Plant. e-mail: subscriptions@nature.com

the urchins fed. This study, along with the work of others⁷⁻⁹, points the way to dissecting out the details of the causes of mass extinctions — the search for correlations (or lack thereof) within and between detailed palaeontological data sets. □

Charles R. Marshall is in the Department of Earth and Space Science, Molecular Biology Institute and Institute for Geophysics and Planetary Physics, University of California, Los Angeles, California 90095-1567, USA.

e-mail: marshall@ess.ucla.edu

1. Smith, A. B. & Jeffery, C. H. *Nature* 392, 69–71 (1998).
2. Raup, D. M. *Paleobiology* 17, 37–48 (1991).

3. Sepkoski, J. J. & Kendrick, D. C. *Paleobiology* 19, 168–184 (1993).
4. Roman, J. *Bull. Sect. Sci.* 6, 133–147 (1984).
5. Marshall, C. R. & Ward, P. D. *Science* 274, 1360–1363 (1996).
6. Jeffery, C. H. & Smith, A. B. in *Echinoderms: San Francisco* (eds Mooi, R. & Telford, M.) (Balkema, Rotterdam, 1998).
7. Raup, D. M. & Jablonski, D. *Science* 260, 971–973 (1993).
8. Jablonski, D. & Raup, D. M. *Science* 268, 389–391 (1995).
9. Jablonski, D. in *Extinction Rates* (eds Lawton, J. H. & May, R. M.) 25–44 (Oxford Univ. Press, 1995).
10. Schultz, P. H. & D'Hondt, S. *Geology* 24, 963–967 (1996).
11. Paul, C. R. C. & Mitchell, S. F. *Geology* 22, 679–682 (1994).
12. D'Hondt, S., Herbert, T. D., King, J. & Gibson, C. *Geol. Soc. Am. Spec. Pap.* 307, 303–317 (1996).
13. Sepkoski, J. J. in *Global Events and Event Stratigraphy* (ed. Walliser, O. H.) 35–51 (Springer, Berlin, 1995).
14. Jeffery, C. H. *Palaentology* 40, 659–712 (1997).

Cardiac physiology

A last wave from the dying heart

Arun V. Holden

Throughout our lives the heart beats rhythmically — synchronized contraction of the ventricles pumps blood through the body. But during the last few minutes of life (in almost all non-violent deaths), contraction of the ventricular muscle is no longer synchronized. Instead, the ventricles writhe and fibrillate at a frequency some ten times faster than that of the heart rate. If the fibrillating heart is held in your hand it feels like a ball of writhing worms, quivering chaotically. It no longer ejects blood, and death is imminent.

On pages 75 and 78 of this issue, Gray *et al.*¹ and Witkowski *et al.*² describe the spatio-temporal pattern of electrical activity on the surface of isolated, blood-perfused dog, rabbit and sheep hearts during ventricular fibrillation. The fibrillation is induced by electrical stimulation, and the membrane potential of surface cells is monitored using voltage-sensitive dyes and optical imaging techniques. The resulting movies show the onset and development of electrical activity during fibrillation, with a resolution of about 0.5 mm and 1.2 ms (see Supplementary information, ref. 3). And these remarkable images indicate that, rather than being a chaotic process, cardiac fibrillation involves an unexpected amount of spatial and temporal patterning and order.

The contraction of cardiac muscle cells is triggered by electrical excitation. Each beat is driven by a wave of excitation, spreading out from the pacemaker region of the heart — the sino-atrial node, which is autorhythmic — and through the entire organ. If this pattern of propagation breaks down, fibrillation may result. During fibrillation, an irregular component appears in electrocardiogram recordings at about 8–10 Hz. This indicates that there is temporal order in the electrical activity that produces fibrillation⁴, consistent with some remaining spatio-temporal patterning or order in the activity of the fibrillating heart. Fibrillation is thought to result from

so-called re-entrant waves of excitation, which rotate around a core (as a spiral wave in two dimensions) or a filament (as a scroll in three dimensions). The electrophysiological structures that emit such rotating waves — rotors — arise in excitable models of cardiac tissue^{5,6}. But although such ordered propagation (as spiral or scroll waves) is found in computational models of the heart⁷, it has not been demonstrated *in situ*.

It should be simple to detect and quantify a spiral wave on the surface of the heart. However, this is complicated by the three-dimensional geometry of the heart, and the differences in conduction velocity throughout the organ. Moreover, a simple rotor need not manifest itself on the surface of the heart as an ideal spiral. Witkowski *et al.*² approached the problem by evaluating the spatio-temporal correlation function. This

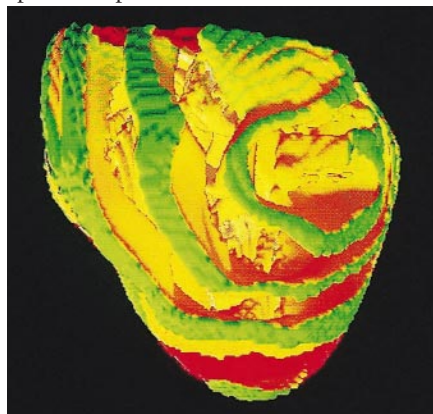


Figure 1 Propagation of a simple re-entrant wave in an anatomically detailed model of the canine ventricle. The geometry of the cardiac muscle fibres determines anisotropy. Gray *et al.*¹ and Witkowski *et al.*² have studied the propagation of re-entrant waves during cardiac fibrillation and find that, rather than being a completely chaotic process, there is a high degree of spatial and temporal order. (Anatomical details from ref. 13; computation and visualization by George Kremmydas.)

measures how activity at a point is correlated with previous activity elsewhere. The authors identified a high, almost periodic, correlation function with the order produced by rotors on the surface of the heart. During the onset of fibrillation they found a period of high order but, after a few minutes of fibrillation, the activity on the surface was no longer as ordered.

Activity at any point on the surface of the heart during propagation of a re-entrant wave has a strong periodicity, which is reflected in the power spectrum of the electrocardiogram during clinical fibrillation⁴. So Gray *et al.*¹ plotted the spatial activity in terms of the phase of oscillation, rather than voltage. Their approach allows the core of a rotor to be identified as a phase singularity, around which phase changes through an entire cycle. The evolution of fibrillation may be followed by tracking the birth, death and motion of these phase singularities, providing a different way to track the changes that occur in the fibrillating heart.

Both of these papers^{1,2} emphasize the spatial patterning and order that persist during fibrillation, in contrast to the irregular activity at points, which can be considered as chaotic processes that are amenable to chaos-control techniques⁸. The order results from the properties of propagating waves in excitable media. Defibrillation techniques can be separated in terms of their effects on propagation of the wave-front⁹ or on movement of the rotor¹⁰. And it is during the early stages of fibrillation that methods influencing rotor motion may be effective.

In the ventricle, waves are propagated more rapidly along the orientation of the muscle fibres, within a complicated and constantly moving geometry (the fibrillating heart). These three-dimensional effects radically influence the patterns that are seen on the surface of the heart. Rather than inferring (from surface patterns) or modelling the three-dimensional patterns of activity during fibrillation⁷, experimental means of reconstructing and visualizing the patterns of three-dimensional spatio-temporal activity within the heart are needed. Optical tomography has been used to visualize three-dimensional waves in chemical excitable media¹¹. This, combined with voltage-sensitive dyes, may allow the direct reconstruction of three-dimensional propagation in small hearts. Such hearts are almost transparent to intense illumination by an appropriate wavelength.

The high-resolution techniques described by Gray *et al.*¹ and Witkowski *et al.*² provide the basis for directly investigating the effects of applied electric fields on the heart. We should be able to assess, experimentally, theoretical concepts based on nonlinear wave mechanisms¹². These techniques may also help us to understand the spatio-temporal effects of applied currents, and the proposed defibrillation and pacemaking techniques.

Perhaps the methods used to control re-entrant waves in excitable media¹⁰ may also be applied to control fibrillation in human patients. □

Arun V. Holden is at the Centre for Nonlinear Studies, University of Leeds, Leeds LS2 9JT, UK.
e-mail: arun@cbiol.leeds.ac.uk

1. Gray, R. A., Pertsov, A. M. & Jalife, J. *Nature* **392**, 75–78 (1998).
2. Witkowski, F. X. *et al.* *Nature* **392**, 78–82 (1998).
3. <http://www.nature.com>
4. Clayton, R. H., Murray, A. & Campbell, R. W. F. *J. Cardiovasc. Electrophysiol.* **6**, 616–624 (1995).
5. Holden, A. V. *Nature* **387**, 655–666 (1997).

Cavitation physics

Lasers blow a bigger bubble

Detlef Lohse

Snorkellers enjoy listening to drops falling on the surface of the sea. Most of the sound is not generated when the drops hit the surface, but when the little voids in the water that form after the collision collapse¹. Kinetic energy is transferred into sound, as well as into thermal energy. But what about light? The typical kinetic energy of a falling molecule in a drop is only 10^{-6} eV, whereas the typical energy of a photon of light is in the eV range, so we would need an energy concentration of six orders of magnitude to produce light.

The voids formed by falling raindrops may not collapse violently enough to emit light, but Ohl, Lindau and Lauterborn² of the University of Göttingen were able to detect light on collapse of an artificially created void. By firing an eight-nanosecond infrared laser pulse (containing just 20 mJ of energy) into water, they can cause a single bubble to nucleate at the laser focus. The bubble fills with dissolved gas and water vapour and grows up to three millimetres in diameter — with a volume about 100–1,000 times larger than its equilibrium value. As the pressure inside the bubble is much less than that outside, the ambient pressure makes the bubble implode. At collapse, the authors detect up to 3×10^7 photons.

The authors call this phenomenon single-cavitation bubble luminescence (SCBL). It had already been observed in the 1970s by Russian groups³, but no quantitative measurements of the light emission could then be made. SCBL resembles single-bubble sonoluminescence (SBSL), a peculiar phenomenon discovered^{4,5} in 1990. In SBSL a single bubble is trapped in an oscillating sound field and repeatedly collapses and emits light^{4–6}.

Neither in SCBL nor in SBSL are the details of the light production process understood. What is already clear in both phenomena is that, on collapse, the gas inside the bubble is compressed so that it heats up, as in a pumped-up bicycle tyre. Here, however, it is heated rapidly to a temperature of at least sev-

6. Biktashev, V. N. & Holden, A. V. *Chaos* **8**, 48–56 (1998).
7. Keener, J. & Panfilov, A. V. in *Cardiac Electrophysiology — From Cell to Bedside* 2nd edn (eds Zipes, D. & Jalife, J.) 335–347 (Saunders, Philadelphia, 1995).
8. Garfinkel, A. *et al.* *J. Clin. Invest.* **99**, 305–314 (1997).
9. Pumir, A. & Krinsky, V. I. *Physica-D* **91**, 205–219 (1996).
10. Biktashev, V. N. & Holden, A. V. *Chaos Solitons Fractals* **5**, 575–622 (1995).
11. Winfree, A. T., Caudle, S., Chen, G., McGuire, P. & Szilagyi, Z. *Chaos* **6**, 617–626 (1996).
12. Biktashev, V. N. in *Computational Biology of the Heart* (eds Panfilov, A. V. & Holden, A. V.) 137–171 (Wiley, Chichester, 1997).
13. Nielsen, P. M. F., Le Grice, L. J., Smail, B. H. & Hunter, P. J. *Am. J. Physiol.* **260**, H1365–H1378 (1991).

eral thousand kelvin, so visible light can be emitted. The three most likely light-emission mechanisms are ordinary thermal black-body radiation, collision-induced emission, and bremsstrahlung.

The new method may allow measurements that shed some light on the emission mechanism, as the created bubbles are up to 1,000 times larger in volume than those of SBSL. SCBL also offers the opportunity to study bubble–bubble interactions, because two bubbles can be nucleated with controlled relative distance and phase. A better knowledge of bubble–bubble interactions will be the first step towards understanding multi-bubble sonoluminescence (MBSL), the light emission by thousands of bubbles in a bubble cloud under oscillating high pressure. Here, the bubbles disturb one another, so the collapses are much weaker on average, and the spectra of the emitted light have distinct lines⁷, in contrast⁶ to those of SBSL — implying that the gas in SBSL is so hot that emission lines are thermally smeared out. Ohl *et al.*³ have already taken a step towards studying bubble interactions by generating a bubble near a wall. As expected, the symmetry breaking due to the wall made the collapse less spherical, resulting in less light.

MBSL is of particular interest to chemists, as the kinetic energy of collapsing bubbles can be transformed not only into light but also into chemical reactions. In an experiment reported last year, Suslick *et al.*⁸ measured reaction rates in a 200-m-s⁻¹ water jet. This strongly turbulent flow creates tiny bubbles, which collapse, heating the trapped gases and the water vapour and so partly dissociating them. The radicals formed can recombine to oxidative species such as OH* and H₂O₂, which oxidate dissolved iodide to iodine — easily detectable, as it forms a blue complex with starch. Although it was already known that such reactions can occur in acoustical cavitation (sonochemistry⁹), it is remarkable that — provided some starch is dissolved — extremely strong stirring of a 1-molar potassi-

um iodide solution may turn it blue.

Suslick *et al.* also discovered a counter-intuitive water-temperature dependence of the reaction rate: as the water temperature increases the reaction rate decreases. The reason, presumably, is that the water vapour pressure strongly increases with temperature, and on fast compression a water-vapour-filled bubble does not heat up as easily as a nitrogen bubble, or even an inert gas bubble.

The common feature of all the phenomena discussed so far — the sound of rain, SCBL, chemistry-inducing hydrodynamical cavitation, MBSL and SBSL (in which chemical reactions also occur^{10,11}) — is that a bubble collapses, forming a geometric singularity at its centre where extreme physical conditions are achieved. The main questions now are, how extreme can these conditions be, and what physical processes occur under such conditions?

Great progress towards answering these questions has been made by Gompf, Eisenmenger and co-workers¹². With time-correlated single-photon counting, they were able to measure the width of SBSL light pulses. Their measurements, which have now been confirmed in other labs, show that the light can last as long as 300 ps. Moreover, the length of the light pulse is approximately the same for all observable frequencies, constraining theoretical models of the light production mechanism. For example, if the emission is simply black-body radiation, the temperature must reach at least 70,000 K.

It may be worth trying to increase the intensity of the bubble collapse by combining the methods employed in SCBL and SBSL. One could nucleate a bubble using a strong laser pulse in the pressure antinode of an oscillating sound field, so that the increasing sound pressure would accelerate the bubble collapse. In this way the temperatures and pressures inside the bubble could be pushed even further into the region of extreme states of matter, and time-resolved spectroscopy of a single light pulse might become possible. □

Detlef Lohse is in the Department of Physics, University of Marburg, Renthof 6, D-35032 Marburg, Germany.

e-mail: lohse@stat.physik.uni-marburg.de

1. Prosperetti, A., Crum, L. A. & Pumphrey, H. C. *J. Geophys. Res.* **94**, 3255–3259 (1989).
2. Ohl, C. D., Lindau, O. & Lauterborn, W. *Phys. Rev. Lett.* **80**, 393–396 (1998).
3. Buzukov, A. A. & Teslenko, V. S. *JETP Lett.* **14**, 189–191 (1971).
4. Gaitan, D. F. Thesis, Univ. Mississippi (1990).
5. Crum, L. A. *Phys. Today* **47**, 22–29 (1994).
6. Barber, B. P., Hiller, R. A., Löfstedt, R., Putterman, S. J. & Weninger, K. R. *Phys. Rep.* **281**, 65–144 (1997).
7. Matula, T. J., Roy, R. A., Mourad, P. D., McNamara, W. B. & Suslick, K. S. *Phys. Rev. Lett.* **75**, 2602–2605 (1995).
8. Suslick, K. S., Mdeleni, M. M. & Ries, J. T. *J. Am. Chem. Soc.* **119**, 9303–9304 (1997).
9. Suslick, K. S. *Science* **247**, 1439–1445 (1990).
10. Lohse, D., Brenner, M. P., Dupont, T., Hilgenfeldt, S. & Johnston, B. *Phys. Rev. Lett.* **78**, 1359–1362 (1997).
11. Matula, T. J. & Crum, L. A. *Phys. Rev. Lett.* **80**, 865–868 (1998).
12. Gompf, B., Günther, R., Nick, G., Pecha, R. & Eisenmenger, W. *Phys. Rev. Lett.* **79**, 1405–1408 (1997).

Neurodegeneration

Chaperoning brain diseases

William J. Welch and Pierluigi Gambetti

A common feature of several neurodegenerative disorders — including Alzheimer's and Parkinson's diseases — is a pathogenetic mechanism which is similar to that proposed for prion diseases (Table 1). The proteins involved have poorly structured native conformations that can be destabilized further by genetic mutations, leading them to adopt β -sheet structures. These, in turn, result in the formation of insoluble, disease-causing protein aggregates in the brain. Under some conditions the aggregates can then act as a seed, inducing the normal protein to adopt the abnormal conformation^{1,2}. But might these apparent 'errors in protein folding' be mediated by other macromolecules? This obvious but unresolved issue is addressed by two papers^{3,4} from Lindquist and colleagues in *Proceedings of the National Academy of Sciences*.

The idea that a molecular chaperone might participate in the conversion process seems plausible and has, in fact, been suggested for the prion-related disorders. Here, the apparent 'species barrier' (where one species of animal is resistant to infection by prions derived from another species) has led to the conclusion that another component must feature in the conversion of the normal prion protein, PrP^C, into its infectious and pathological form, PrP^{res}. Indeed, Prusiner and colleagues² have proposed that the conversion of PrP^C to PrP^{res} is facilitated by a host component, referred to as protein X.

Lindquist and colleagues have examined the interactions between various molecular chaperones and two prions — yeast Sup35 and the mammalian prion protein. Sup35 is a translation termination factor that can exist in a soluble [psi⁺] form, or an aggregated [PSI⁺] form which assembles into fibrils resembling the prion protein in animals. Schirmer and Lindquist³ used circular dichroism to show a selective interaction between Sup35 and heat-shock protein (hsp)104. (Genetic studies have already shown that this chaperone can influence the physical state of Sup35 *in vivo*.) Over time, the hsp104–Sup35 mixture showed an increase in light scattering, indicating that one or both proteins were undergoing some type of conformational change — most likely aggregation.

DeBburman *et al.*⁴ showed that yeast hsp104 and a second molecular chaperone, the bacterial GroEL protein (without its normal *in vivo* cofactor GroES), can interact with the mammalian prion protein. Moreover, both proteins promoted an increase in

the efficiency of the PrP^C to PrP^{res} conversion. But, as the authors pointed out, both the experimental conditions and the chaperones that were the most effective in the conversion process are not very relevant to the situation *in vivo*. What is important is the fact that the different chaperones have an effect on the PrP^C to PrP^{res} conversion, probably due to their ability to stabilize folding intermediates.

The idea that a third party, such as a molecular chaperone, might be involved in the conversion of proteins from their soluble to insoluble states is intriguing (Fig. 1). Molecular chaperones interact with proteins that are being synthesized, folded or translocated into organelles, as well as with mature proteins that tend to unfold (and so are prone to aggregation) because of environmental insults such as heat shock. Although the different chaperones do not provide any direct information for the folding process, they seem to stabilize intermediates during folding. By doing so, they act to reduce the number of non-productive folding intermediates that might proceed to form intracellular aggregates.

Thus, we normally think of molecular chaperones as reducing protein aggregation inside the cell. But the possibility that a chaperone could feature in the conversion process is not so far-fetched. For example, if a chaperone-bound folding intermediate

(such as a protein rich in β -sheet) is released from the chaperone while in the vicinity of an existing protein aggregate (that is, a 'seed'), the intermediate could conceivably be recruited into the insoluble aggregate. In a similar way, certain isoforms of apolipoprotein E (a risk factor for Alzheimer's disease) are thought to promote or stabilize β -sheet conformations in amyloid- β peptide, thereby possibly facilitating the formation of insoluble amyloid fibres⁵.

Intracellular accumulation of abnormally folded proteins has a significant effect on cells — activation of the heat-shock/stress response. So might neurons harbouring amyloid deposits activate their stress response in an attempt to increase the levels of molecular chaperones? A slight increase in the levels of a cytosolic chaperone (hsp73) has been observed in neuronal cells expressing PrP^{res}, compared with their PrP^C-expressing counterparts⁶. Perhaps more interesting was the fact that cells expressing PrP^{res}, in contrast to cells making only PrP^C, could not activate the heat-shock/stress response.

Why might this be? Perhaps the heat-shock response is activated as PrP^{res} begins to accumulate. If, over time, however, the chaperones cannot re-fold or remove PrP^{res}, the cells may then 'turn off' their heat-shock response. Or perhaps infection results in a type of clonal selection — cells with an already attenuated heat-shock response are the only ones able to propagate PrP^{res}. If that is the case, one wonders whether an attenuated stress response might also be common to neuronal cells accumulating other abnormally folded proteins, such as those involved

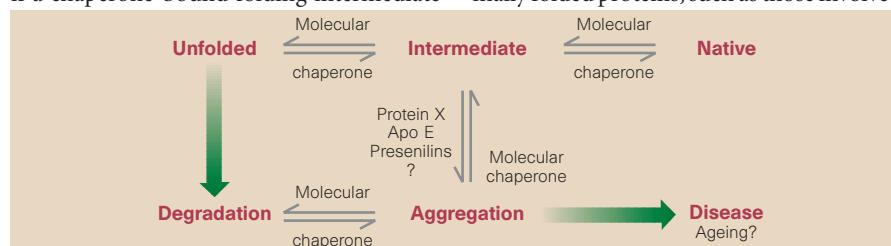


Figure 1 Relationships between protein folding, neurodegenerative diseases and ageing. Lindquist and colleagues^{3,4} have observed *in vitro* interactions between prion proteins and molecular chaperones. Might chaperones — either classical molecular chaperones or, perhaps, substrate-specific chaperones such as protein X, apolipoprotein E or presenilins — feature in the conversion of prions, from their soluble to insoluble forms, through the stabilization of a certain folding intermediate or intermediates?

Table 1 'Aggregate-forming' neurodegenerative diseases

Disease	Protein	Structure*	Aggregate	Location
Prion ¹⁰	Prion	α -helix and random coil	β -pleated, protease resistant and amyloid	Extracellular
Alzheimer's ^{5,8}	Amyloid- β	α -helix and random coil	β -pleated and amyloid	Extracellular
Parkinson's ^{12–14}	α -synuclein	Imperfect repeats	Lewy body, insoluble (β -pleated?)	Cytoplasmic
Huntington's ⁵	Huntingtin	Trinucleotide repeats	Insoluble (β -pleated?)	Nuclear (intracytoplasmic?)
Spinocerebellar ataxia 1 and 3 (ref. 16)	Ataxin 1 and 3	Trinucleotide repeats	Insoluble (β -pleated?)	Nuclear

*Refers to the dominant primary and secondary structure.

in the different neurodegenerative diseases. Moreover, because the stress response helps to protect cells against a variety of metabolic insults, could failure of the cells to activate the stress response lead to increased neuronal vulnerability in response to other traumas?

The accumulation of abnormally folded proteins is also a hallmark of ageing. Moreover, as cells age they lose their ability to activate the stress response^{7,8}. So might the neurodegenerative diseases that involve the accumulation of abnormally folded proteins be a manifestation of ageing, with mutations in particular gene products (Table 1) accelerating the process? Orgel presented⁹ a similar argument 35 years ago. In a strictly theoretical paper, he suggested that the accumulation of protein-folding errors might help to explain the phenomenon of ageing. In his 'error-catastrophe' model, mistakes in protein folding (especially of proteins involved in protein-synthetic pathways) might initiate a type of self-perpetuation. As more mistakes occurred in protein folding, these errors would lead to further errors such that, eventually, the cell would find itself awash in abnormally folded proteins. Perhaps even more prophetic was his suggestion that "special proteins might be synthesized which are converted by a certain class of errors into lethal polypeptides". Could the proteins associated with neurodegenerative disorders represent a type of 'lethal polypeptide'?

All of the information about the molecular basis of what were thought to be different neurodegenerative diseases seems to be

culminating into a central theme — the conversion of soluble proteins into insoluble aggregates which, over time, give rise to neurodegenerative pathology. How these protein aggregates affect neuronal function, eventually leading to neurodegeneration, are just some of the issues that now need to be addressed.

William J. Welch is in the Department of Surgery, Medicine and Physiology, University of California at San Francisco, San Francisco, California 94143, USA.

e-mail: welch@itsa.ucsf.edu

Pierluigi Gambetti is in the Division of Neuropathology, Institute of Pathology, Case Western Reserve University, 2085 Adelbert Road, Cleveland, Ohio 44106, USA.

e-mail: pxg13@po.cwru.edu

1. Ghiso, J., Wisniewski, T. & Frangione, B. *Mol. Neurobiol.* **8**, 49–64 (1994).
2. Paulson, H. L. *et al. Neuron* **19**, 333–334 (1997).
3. Schirmer, E. C. & Lindquist, S. *Proc. Natl Acad. Sci. USA* **94**, 13932–13937 (1997).
4. DeBburman, S. K. *et al. Proc. Natl Acad. Sci. USA* **94**, 13938–13943 (1997).
5. Wisniewski, T., Ghiso, J. & Frangione, B. *Neurobiol. Dis.* **4**, 313–328 (1997).
6. Tatzelt, J. *et al. Proc. Natl Acad. Sci. USA* **92**, 2944–2948 (1995).
7. Liu, Y. C. *et al. J. Biol. Chem.* **264**, 12037–12043 (1989).
8. Fargnoli, J. *Proc. Natl Acad. Sci. USA* **87**, 846–850 (1990).
9. Orgel, L. E. *Proc. Natl Acad. Sci. USA* **49**, 517–521 (1963).
10. Prusiner, S. B. *Science* **278**, 245–251 (1997).
11. Frangione, B., Wisniewski, T., Castanol, E. M. & Ghiso, J. in *Research in Alzheimer's Disease and Related Disorders* (eds Iqbal, K., Mortimer, J. A., Winblad, B. & Wisniewski, H. M.) 563–568 (Wiley, New York, 1995).
12. Polymeropoulos, M. H. *et al. Science* **276**, 2045–2047 (1997).
13. Goedert, M. *Nature* **388**, 232–234 (1997).
14. Nussbaum, R. L. & Polymeropoulos, M. H. *Hum. Mol. Genet.* **6**, 1687–1691 (1997).
15. Heitz, N. & Zoghbi, H. *Nature Genet.* **16**, 325–327 (1997).
16. Ross, C. A. *Neuron* **19**, 1147–1150 (1997).

Cosmology

A circumscribed Universe

Masataka Fukugita

Our Universe occupies space. But what sort of space? Mathematicians call different types of space 'manifolds', and characterize them by two properties: by the local curvature — geometry; and the overall shape — topology. Most cosmologists are concerned only with the curvature, as it is directly related to the mass in the Universe (and therefore it is observable), and because it determines the past and future evolution of the Universe with time. The global shape is not, in general, prescribed by geometry, but it also is observable: topological information is imprinted on the fluctuations of the cosmic microwave background. This has already provided some evidence that the Universe is infinite or, at least, not small compared with the distance we can see. But now Cornish, Spergel and Starkman^{1,2} have realized that the Universe can still be finite, and relatively small, if its curvature is negative. Moreover, we may soon be able to measure its size and shape. This question of

whether the Universe is finite has profound implications for its birth.

Cosmologists classify the curvature of the Universe as positive, zero or negative. Flat, euclidean space, called R^3 , has zero curvature — the sum of internal angles of a triangle drawn on it is 180° . With positive curvature, as on a sphere, the sum of the angles is more than 180° , and with negative it is less. Einstein's equation relates the curvature to the amount of mass present in the Universe (assuming that the energy density of empty space — the cosmological constant — is zero). The sign of the curvature depends on whether the average mass density at present is more than about $2 \times 10^{-29} \text{ g cm}^{-3}$, the critical density.

If the density has exactly this value, the Universe is flat. If it is larger, the Universe has a positive curvature — it is said to be 'closed'. Space is then like the surface of a sphere, and the Universe will eventually collapse to infinite density in a Big Crunch. In this case the Universe must be finite.

If the density is less than critical, the curvature is negative, and the Universe will expand eternally. This manifold is a three-hyperboloid — locally, a three-dimensional analogue of a saddle. This case is referred to as 'open' geometry, with infinite space implicitly in mind. This is, however, a careless term: a negatively curved space can be finite, as can a flat space, if it is multiply connected (see box).

The idea that the Universe may occupy a multiply-connected space is old^{3–6}. The straightforward consequence would be that our Galaxy, or galaxies familiar to us, would be visible in many places, as if we were all in a hall of mirrors. In practice, however, observations are obscured by the evolution of the objects: the images of a single galaxy or cluster will look different, because the light has travelled a different number of times around the Universe and so the images have different ages. This effect is especially strong when the size of space is of the order of the curvature, that is, it takes nearly the age of the Universe for light to travel all the way around.

But the topology of space also affects the fluctuations in the cosmic microwave background, which originates from a time in the early Universe when matter became too thin to trap radiation, at the 'sphere of last scattering'. These fluctuations should be evident on a wide range of angular scales, but finite space would mean that the spectrum of fluctuation sizes should cut off at a wavelength corresponding to the size of the space. Observations^{7–9} by the COBE satellite are consistent with an infinite Universe, or at least one larger than the 'horizon' (the sphere beyond which we cannot see, because light has taken the age of the Universe to reach us from it).

But Cornish and collaborators^{1,2} note that this is not true for a space with negative curvature. In that case, large-angle micro-

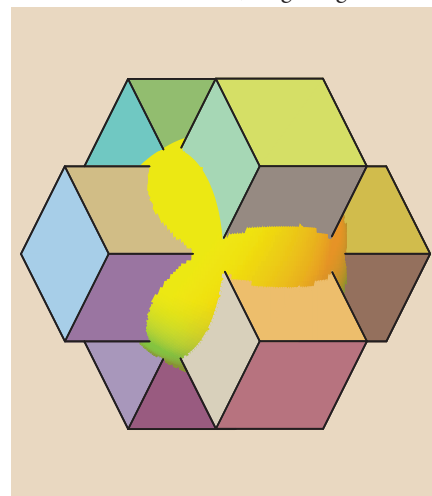
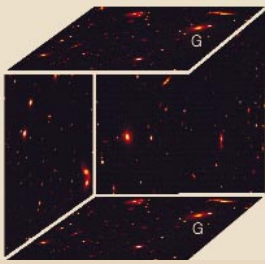


Figure 1 A sphere intersects cubic domains. If the Universe is smaller than the sphere of last scattering (where the cosmic microwave background originates), these intersections should be visible as circles of identical microwave background fluctuation on the sky.

GLENN STARKMAN

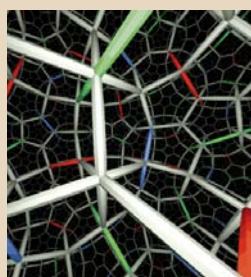
Small 'open' spaces

How can a space be finite and yet have no edges? The simplest way to achieve it is to take a cube, and identify each face with that on the opposite side. In the diagram below, galaxy G is on both the top and bottom of the cube – this is called a periodic boundary condition, and makes the space multiply connected. Travelling in any straight line, you would cross the cube repeatedly. This cube is known as



the fundamental domain of the space. Its boundaries are arbitrary: there is no physical threshold to cross, and, looking into the distance, you would see patterns and objects repeating themselves – including the back of your own head. The space thus defined is called a three-torus, T^3 , and is locally identical to ordinary euclidean space, R^3 .

Similarly, there are finite spaces with negative curvature, or 'compact three-hyperboloids'. Examples can be constructed by identifying opposite faces of a suitable polyhedron. For example, a dodecahedron with opposite faces fitted



together after rotating by 108° spans a three-hyperboloid space, known as the Seifert-Weber space (top right¹²). There are infinitely many types of three-hyperboloid^{13,14}. Their volumes are related to their curvature, unlike the torus in flat space whose volume is arbitrary. The volume is at least larger than $0.00082 r^3$, where r is the radius of curvature. **M.F.**

wave-background fluctuations are dominated by the gravitational potential at low redshifts (in the nearby Universe), and are hardly anything to do with physics at the last scattering sphere. So if the Universe has less than critical density, the COBE observation does not give a strong constraint on the size of the Universe.

Instead, the size of our space should be betrayed by identical circles of fluctuation on the microwave sky, where the horizon-size sphere of last scattering crosses the boundary of the fundamental domain centred on Earth (see box). This is conceived best by considering the intersection of a sphere with a slightly smaller cube. The intersection on each face of the cube forms a circle, and because of the periodic boundary condition, a circle on the upper face, say, must be identical to that on the lower face. The number, the relative angle and the size of identical circles will tell us what topology our three-hyperboloid has. This is no distant dream. Two satellites in the planning stages will be able to make such measurements: NASA's microwave-background mission MAP, planned for launch in the year 2000; and ESA's PLANCK, scheduled for 2006.

Whether the curvature of the Universe is negative or flat is still a matter of debate (no evidence supports positive). But the majority of recent observations — of the expansion speed versus the age of the Universe, of the evolution of galaxy clustering, of the brightness of distant supernovae — are in favour

of a low-density Universe. If empty space has no inherent energy (another matter of debate), low density means negative curvature. A precise measurement of the microwave fluctuations on a small angular scale could be definitive, so MAP and PLANCK may be able to answer this part of the question too.

So what is the physical significance of

topology? Some theorists speculate that the Universe was created from nothing^{10,11}, as a quantum vacuum fluctuation. That idea requires the Universe to be finite. Another theoretical advantage of a compact three-hyperboloid is efficient chaotic mixing². Any excitation created in this manifold dissipates rapidly: excitation of one mode excites all other modes, and this smears out inhomogeneities that would have existed in the pre-inflation epoch, solving a problem that faces inflationary cosmologists. In a negative-curvature Universe, inflation takes place only partially and does not produce a Universe as homogeneous as the one we observe, unless we start with a homogeneous initial condition. Chaotic mixing could prepare such an initial state. Perhaps the Universe is so smooth because it is small. $\square$

Masataka Fukugita is at the Institute for Cosmic Ray Research, University of Tokyo, Tanashi, Tokyo 188, Japan, and the Institute for Advanced Study, Princeton, New Jersey 08540, USA.
e-mail: fukugita@sns.ias.edu

1. Cornish, N. J., Spergel, D. N. & Starkman, G. D. *Proc. Natl. Acad. Sci. USA* **95**, 82–84 (1998).
2. Cornish, N. J., Spergel, D. N. & Starkman, G. D. *Phys. Rev. Lett.* **77**, 215–218 (1996).
3. Friedmann, A. Z. *Phys.* **21**, 326–332 (1924).
4. Ellis, G. F. R. *Gen. Rel. Grav.* **2**, 7–21 (1971).
5. Sokoloff, D. D. & Shvartsman, V. F. *Sov. Phys. JETP* **39**, 196–200 (1974).
6. Gott, J. R. *Mon. Not. R. Astron. Soc.* **193**, 153–169 (1980).
7. Sokoloff, I. Yu. *JETP Lett.* **57**, 617–621 (1993).
8. Starobinsky, A. A. *JETP Lett.* **57**, 622–625 (1993).
9. de Oliveira-Costa, A. & Smoot, G. F. *Astrophys. J.* **448**, 477–481 (1995).
10. Zel'dovich, Ya. B. & Grishchuk, L. P. *Mon. Not. R. Astron. Soc.* **207**, 239–289 (1984).
11. Hawking, S. W. *Nucl. Phys. B* **239**, 257–276 (1984).
12. http://www.geom.umn.edu/graphics/pix/Special_Topics/Hyperbolic_Geometry/HSpace2.html
13. Thurston, W. P. & Weeks, J. R. *Sci. Am.* July, 108–120 (1984).
14. Thurston, W. P. & Levy, S. *Three-Dimensional Geometry and Topology* (Princeton Univ. Press, 1997).

Systematics

Sequences lead to tree of worms

Claus Nielsen

Nematode worms are found just about everywhere, often in enormous numbers. Parasitic species live in almost all animals and plants, and only arid soils and the open oceans seem to be unsuitable for free-living species. Yet only about 15,000 species have been formally described, and some textbooks treat nematodes as one of the 'minor phyla'. In spite of this, it is now believed that the number of living species should be counted by the million. Classification has been difficult because most of the species are small to microscopic in size, and they lack obvious distinguishing characteristics. But the first attempt at a phylogenetic classification — based on small-subunit ribosomal DNA sequences from 53 species — is presented by Blaxter *et al.* on page 71 of this issue¹. And, according to their findings,

only one of the two classes of nematode that are recognized in the traditional classification is natural, consisting of an ancestral species and all its descendants.

Free-living nematodes are found in terrestrial soils and marine sediments, where they decompose plant and animal material. Parasitic species of nematode infect common crops such as potatoes, soya beans and corn, as well as livestock, including pigs, cattle and chicken. Human parasites include species that cause crippling or fatal diseases such as filariasis (elephantiasis), trichinosis and hookworm disease. Conversely, parasitic species of nematode are increasingly being used in biological control — some species directly attack and destroy the larvae or adults of plant parasitic insects, whereas others transmit bacteria that destroy the parasite (Fig. 1). $\blacktriangleright$

Most (but not all) nematodes are small and nondescript. For example, *Placentonema gigantissima*, which lives as a parasite in the placenta of sperm whales, grows to a length of 8 m, with a diameter of 2.5 cm. The free-living, marine *Draconema* has elongate adhesive organs on the head and along the tail, and moves like a caterpillar. But the general uniformity of most nematode species has hampered the establishment of a classification that includes both free-living and parasitic species. Two classes have been recognized (the Secernentea and Adenophorea), based on the presence or absence of a caudal sense organ, respectively. But Blaxter *et al.*¹ have concluded from the DNA sequences that the Secernentea is a natural group *within* the Adenophorea. Based on studies of free-living species, a paraphyletic nature for the Adenophorea — that is, a group comprising an ancestor but not all of its descendants — has previously been suggested (for example, by Lorenzen²), but the position of the various parasitic groups has always caused trouble.

One of the most interesting results of the new phylogeny is the discovery that there have been many parallel shifts of feeding strategy within the phylum. The ancestral form was obviously free-living, but the results of Blaxter *et al.* support the idea that parasitism has evolved independently many times. Of the plant parasites, for example, the order Triplonchida comprises only plant parasites, whereas Dorylaimida comprises both omnivores and plant parasites. And the sister orders Aphelenchida and Tylenchida both comprise fungivores, plant parasites — among which the eelworms (Fig. 2) parasitize many important crops, such as potato and sugar beet — and animal parasites.

The animal parasites also belong to several groups that probably evolved independently. Outside the Secernentea, the Tricocephalida comprises mammalian parasites (such as the trichina worm) which do not have an intermediate host. But the Mermithida mainly comprises species that have a juvenile stage in which they infest insects. Within the Secernentea are the Strongylida



Figure 1 The good — *Steinernema bibionis*, a free-swimming parasitic nematode used for biological control of vine weevils.



Figure 2 The bad — eelworm (root knot nematode), which forms characteristic nodules on the roots of sugar beet and rice.

and Rhabditoidea (which are probably sister groups), Strongyloididae, Apelenchida and Tylenchida. The Strongylida comprises vertebrate parasites without an intermediate host (an example is the hook worm). The Rhabditoidea, by contrast, comprises free-living species, such as the favourite experimental model *Caenorhabditis elegans*, and insect parasites. And the Strongyloididae comprises mammalian parasites, many of which infest horses, pigs and cattle.

Blaxter *et al.* also identified a large parasitic group within the Secernentea, comprising three groups of vertebrate parasites (the Ascaridida, Spirurida and Oxyurida) and one group of invertebrate parasites (the Rhigonematida). Of the Ascaridida, some species (for example, *Ascaris* spp.) live in vertebrate intestines. But others, such as species of *Anisakis*, have more complex life cycles, with a crustacean as the first host, a fish as the second host and a fish-eating bird or mammal as the final host. The Spirurida live in vertebrate tissues, and they are transferred by biting or sucking insects. Well-known species include the filaria worm and the whip worm. Of the Oxyurida, the pinworm is a common but harmless human parasite. Finally, the Rhigonematida comprises parasites of terrestrial arthropods.

The establishment of a natural — that is, phylogenetic — classification for the nematodes is an absolute necessity for all aspects of nematode studies, practical as well as theoretical. Moreover, the classification described by Blaxter *et al.* will be an invaluable tool for parasitologists, who search for relationships between parasitic species. They can use this information to look for free-living relatives of important parasites that may be difficult to culture, or for ways in which to combat pests. □

Claus Nielsen is at the Zoological Museum, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark.

e-mail: cnielsen@zmuc.ku.dk

1. Blaxter, M. L. *et al.* *Nature* **392**, 71–75 (1998).

2. Lorenzen, S. in *Concepts in Nematode Systematics* (eds Stone, A. R., Platt, H. M. & Khalil, L. F.) 11–23 (Academic, New York, 1983).

Daedalus

Fibre couplings

The nail, says Daedalus, is a brilliant and versatile fastener, but with a fundamental contradiction. While being hammered in, it is a strut, loaded in compression. It must be thick enough to resist buckling. Yet once in place it is a tie, loaded in tension, and should be thin and flexible to bear its load efficiently. He is now resolving this contradiction.

An ideal nail, he says, should be driven in by a force applied, not to its head, but to its point. Its shaft would then be drawn in under tension; it could not buckle, and would form a perfect tie. But how to apply a force to the point of a nail? Well, suppose the blow on its head lasted only a microsecond. In this time, the shock would travel only a millimetre or so down the nail. The compressed region would be far too short to buckle. The pulse would travel down the nail, and would force the point into the material being fastened. Any reflected energy would travel safely back up the nail as a tension.

Now only the most rigid materials could have an impact time of a microsecond. A diamond hammer driving a diamond nail would be wonderful engineering but disastrous economics. But in electronics, microseconds are positively leisurely. A piezoelectric transducer could hit a nail thousands of times a second. Quartz is piezoelectric, and quartz fibres have amazing tensile strength. So Daedalus is now inventing the quartz-fibre piezoelectric nail.

His 'piezonail' will be a fine, flexible fibre with plated electrodes, and embedded in a plastic reinforcing jacket. You will fit it into a recess in its pulse-generator 'hammer', hold it firmly against the object to be nailed, and switch on. The burst of pulses will force it silently and instantly into the material, giving a strong, tensioned, firmly bound tie. In thick or hard materials, the piezonail will not even need a head; friction will hold it firmly enough. Many thin ties are superior to one thick one, so the hammer will also accept a 'polynail' containing many parallel piezonails in one jacket.

Construction will be transformed. The global toll of bent nails, bruised thumbs and ringing ears will plummet as the piezonail spreads through engineering, carpentry and DIY bodging. A piezonailed structure will be strong, stable, secure — and somewhat enigmatic. Its myriads of fine fixing fibres will be almost invisible, giving no clue as to what holds it together, or how to get it apart again.

David Jones

Vermeer's vision

Jan Vermeer's mastery of the use of paint was such that we see more in his pictures than is actually there. The painter achieved his illusions by using the picture as a field for perceptual exploration.

Martin Kemp

There are some artists whose works demonstrate the highest levels of intellectual deliberation, and yet for whom we have virtually no first-hand evidence of any framework in pictorial theory or the science of vision.

The absence of substantial written testimony for such 'intellectual' painters as Velázquez, Chardin and, above all, Jan Vermeer could simply indicate that the evidence has been lost. But it is more probably a reflection of a shared conviction that there were no contemporary theories of seeing and representation that could adequately describe the processes of perception and depiction they addressed in their paintings.

It has long been surmised that a number

of seventeenth-century Dutch painters made use of the camera obscura as an aid to their supreme naturalism. Philip Steadman's research on a group of ten paintings by Vermeer set in the same space suggests that the painter set up the far end of the room as a walk-in 'optical chamber'.

Steadman's model of the room confirms the extraordinarily high levels of internal optical consistency in paintings such as *The Music Lesson* – even down to the penumbral fan of shadows beside the mirror on the end wall.

It seems likely that the artist laid down the basic disposition of the forms on the basis of their projection through an aperture in a partition on to a screen attached to the opposite end wall.

What the projected image could not do, however, was to tell Vermeer how to translate the effects into paint in such a way that the resulting image stood as a perceptual analogue of the real scene. Our immediate reaction is to think that Vermeer is a master 'describer', filling his scenes with meticulously rendered detail. But this is not the case. He had learnt, by a hard-won process of pictorial trial and error, that, when the artist wishes to cajole our perceptual system into collaborative action, less is definitely more.

The hues and tones of the generalized patches of paint—virtually abstract on close viewing—are pitched with such deliberative skill within the spatial framework that we irresistibly see more than is actually there.

It is like a very complex version of the optical illusions beloved of psychologists of perception. What Vermeer has discovered is that more compelling illusions can be achieved through encouraging our perceptual system to do the lion's share of the work than through the most niggling assertion of detail.

Yet there remains something uneasy in making such inferences about the artist's ideas from the pictures alone. There is, perhaps, one crumb of comfort. The great microscopist, Antoni van Leeuwenhoek—who somehow managed to see bacteria in a single-lens instrument—was an executor of Vermeer's will. What one would give to hear a conversation between the two great 'see-ers' on the business of human vision! □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.



Jan Vermeer's *The Music Lesson*, c. 1670, in the collection of Her Majesty the Queen, Buckingham Palace, London.



Philip Steadman's model of *The Music Lesson*.

Does acoustic testing strand whales?

Mass strandings of live whales have been explained by proposing many 'natural' or human-related causes¹. I found that a recent stranding of Cuvier's beaked whale coincided closely in time and location with military tests of an acoustic system for submarine detection being carried out by the North Atlantic Treaty Organization (NATO).

Cuvier's beaked whale (*Ziphius cavirostris*) seems to be abundant in the East Ionian Sea (Mediterranean Sea), as indicated by strandings and sightings recorded there from 1992 to 1997. This species is a deep-diving, pelagic cetacean that rarely mass-strands². Only seven strandings of more than four individuals have been recorded since 1963 worldwide, the individuals on these occasions numbering 5, 6, 6, 10, 12, 15 and 19 (refs 3–7). In the Kyparissiakos Gulf specifically, the average number of individual whales stranded every half-year is 0.7 (s.d. = 0.9, $n = 11$), with the exception of a mass stranding that occurred on 12–13 May 1996 (Fig. 1).

From the morning of 12 May until the afternoon of 13 May, we recorded 12 Cuvier's beaked whales stranded alive along the coasts of the Kyparissiakos Gulf. The whales were spread along 38.2 kilometres of coast and were separated by a mean distance of 3.5 km (s.d. = 2.8, $n = 11$). This spread in time and location was atypical, as whales usually mass-strand at the same place and at the same time. Two weeks later, one more animal was found decomposing on a remote beach of the neighbouring Zakynthos Island, 57 km away from the closest stranding on the mainland.

Necropsies of eight stranded animals were carried out, but no apparent abnormalities or wounds were found. Many of the stomach contents that were collected contained cephalopod flesh, indicating that recent feeding had taken place.

After looking for possible causes of the mass stranding, we discovered that 'sound-detecting system trials' had been performed

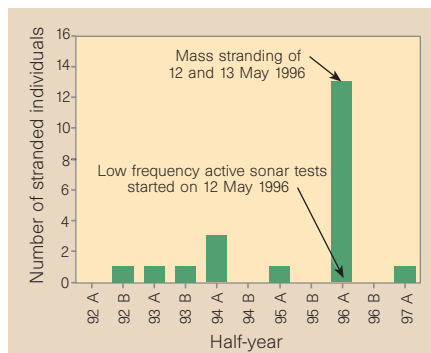


Figure 1 Cuvier's beaked whale strandings in the Kyparissiakos Gulf, 1992–97.

by the NATO research vessel *Alliance* from 24:00 11 May to 24:00 15 May (Warning to mariners 586 of 1996, Hellenic Navy Hydrographic Service) — a period that encompassed the mass stranding. Also, the officially declared area where the sea trials had been carried out enclosed all the coordinates of the stranding points.

The tests that RV *Alliance* performed were for Low Frequency Active Sonar (LFAS), a system for the detection of quiet diesel and nuclear submarines. This system generates extremely loud low-frequency sound (maximum output of ≥ 230 decibels as 1 micropascal broadband waveforms centred at frequencies that range from 250 to 3,000 hertz), which enables long detection ranges^{8,9}.

Research on LFAS began in 1981 and a statement on its environmental impact was formally initiated in July 1996 by the US navy. The adverse effects of low-frequency sound on whales are poorly studied¹⁰, but many specialists warn that at high levels, as occurs with LFAS, they could be dramatic.

The proximity of military manoeuvres has been suspected of causing three previous atypical mass strandings of Cuvier's beaked whales, spread over wide areas of the Canary Islands^{6,11}. On most of the extremely rare occasions that mass strandings are seen in

this species, they show characteristics unlike those that occur with other whales. This suggests that the cause has a large synchronous spatial extent and a sudden onset. Such characteristics are shown by sound in the ocean. Also, deep-diving whales seem to be especially affected by low-frequency sounds, even at quite low received levels^{12,13}.

We know that LFAS was used in the Kyparissiakos Gulf. We also know that no other LFAS tests or mass strandings have occurred in the Greek Ionian Sea since 1981. Taking the past 16.5-year period into account, the probability of a mass stranding occurring for other reasons during the period of the LFAS tests is less than 0.07%.

Although pure coincidence cannot be excluded, it seems improbable that the two events were independent. Little is known about whales' reactions to LFAS; to obtain definitive answers, more information needs to be gathered. But unfortunately, most of the data about the use of LFAS are subject to military secrecy.

A. Frantzis

Zoological Laboratory, Department of Biology,
University of Athens, 15784 Athens, Greece
e-mail: afrantzis@atlas.uoa.gr

- Geraci, J. R. & Lounsbury, V. J. *Marine Mammals Ashore. A Field Guide for Strandings* (Texas A&M Sea Grant Publication, Galveston, Texas, 1993).
- Heyning, J. E. in *Handbook of Marine Mammals* Vol. 4 (eds Ridgway, S. H., Harrison, R. J.) 289–308 (Academic Press, London, 1989).
- Erdman, D. S., Harms, J. & Flores, M. M. *Cetology* 17, 1–14 (1973).
- Robinson, G., Koster, F. & Villa, J. *Noticias de Galapagos* 38, 16–17 (1983).
- Miyazaki, N. *IBI Reports* 1, 21–25 (International Marine Biological Research Institute, 1989).
- Simmonds, M. P. & Lopez-Jurado, L. F. *Nature* 337, 448 (1991).
- Tortorese, E. *Natura* 54, 120–122 (1963).
- Endeavour and Achievement* (NATO–SACLANT Undersea Research Centre, La Spezia, Italy, 1993).
- SACLANT Newsletter* 4 (NATO–SACLANT Undersea Research Centre, La Spezia, Italy, 1996–97).
- Richardson, W. J. & Würsig, B. *Mar. Fresh. Behav. Physiol.* 29, 183–209 (1997).
- Vonk, R. & Martin, V. *European Research on Cetaceans* 3, 73–77 (1989).
- Watkins, W. A., Moore, K. E. & Tyack, P. *Cetology* 49, 1–15 (1985).
- Bowles, A. E. et al. *J. Acoust. Soc. Am.* 96, 2469–2484 (1994).

Electron pairs shed light on frustrated percolation

Cahn, in his News and Views article¹, wondered why an arsenic–boron pair is best suited to 'stuff' a double vacancy in the crystal lattice in preference to a pair of boron atoms or arsenic atoms. The answer can be found if to describe bonding Linnett's model² for the covalent chemical bond is used. This is a generalization of the electron-pair idea which incorporates the Pauli and

Ruedenberg principles and accounts well for molecular structures without resonance or configuration interaction.

Homogeneous dimers of either boron (B) or arsenic (As) in a brass matrix should not be more stable than monomers or homogeneous polymers, and their concentration should be low. Such polymers exist only in the less electronegative metals, so are not expected in zinc, copper or brass.

On the other hand, once boron arsenide (BAs) is formed in a metal, it is unlikely to grow by trapping free B and As atoms alternately — a concept that can be derived from

its Linnett configuration. This way of describing compounds shows that the BAs molecule has three bonds, a pair of non-bonding electrons on the As and an empty site on the B atom. Therefore the molecule still has room to accept charge from the metal matrix. A more complete electron occupation then impedes further growth of the dimer. B and As oligomers exist in metal matrices but the BAs network does not — it exists only in the free cubic BAs solid. The BAs molecule should be stable in a metal matrix and it fits into the zinc double vacancy (divacancy) which has the right size.

Alternative models are worth considering if they suggest new experiments³; this suggests several. First, it should be possible to replace As by P (and perhaps even N) without a diminished anti-corrosive effect in brass. Second, other non-metal pairs as well as B/pnictide, such as C (or Si) with Si (or O), might behave similarly. Third, the solubility of B and As (or P) in brass may depend on each other's concentration and on the divacancy concentration. Novel quaternary intermetallic compounds of B and As might even be found. Finally, the trick should not work in alloys of metals below group 10 which corrode with a divacancy mechanism if formation of metal boride or arsenide phases competes.

P. J. van der Put

Laboratory of Inorganic Chemistry, Delft University of Technology, 2628 BL Delft, The Netherlands
e-mail: p.j.j.m.vanderput@stm.tudelft.nl

1. Cahn, R. W. *Nature* 389, 121 (1997).
2. Linnett, J. W. *The Electronic Structure of Molecules. A New Approach* (Methuen, London, 1964).
3. Oreskes, N., Shrader-Frechette, K. & Belitz, K. *Science* 263, 641 (1994).

Synthetic ligands point to cell surface strategies

Protein shedding, the proteolytic release of a cell surface protein, can serve a regulatory role by liberating soluble molecules into circulation while decreasing their concentration on the cell surface¹. We have created a new class of multivalent ligands, 'neoglycopolymers', which are designed to promote the proteolytic cleavage of a cell adhesion molecule involved in the inflammatory response, L-selectin². These synthetic ligands induce the release of the extracellular portion of L-selectin by appropriating an endogenous protease; such activities suggest new strategies to generate anti-inflammatory agents and regulate the cell surface.

L-selectin mediates the process of leukocyte rolling³, an initial step in the inflammatory response, by recognizing carbohydrate epitopes that are present on endothelial cells. L-selectin is constitutively present on the surface of most leukocytes, but a soluble form is also found in circulation.

The L-selectin ligands that occur physiologically, such as GlyCAM-1 (ref. 4), are mucin-like proteins which contain clusters of O-linked saccharide chains. Such chains could engage L-selectin in multivalent binding at the cell surface. We hypothesized that the clustering of L-selectin, as a consequence of ligand binding, leads to the proteolytic release of its soluble form from the cell⁵.

So we synthesized molecules that, like mucins, present multiple copies of saccharide epitopes on an extended backbone.

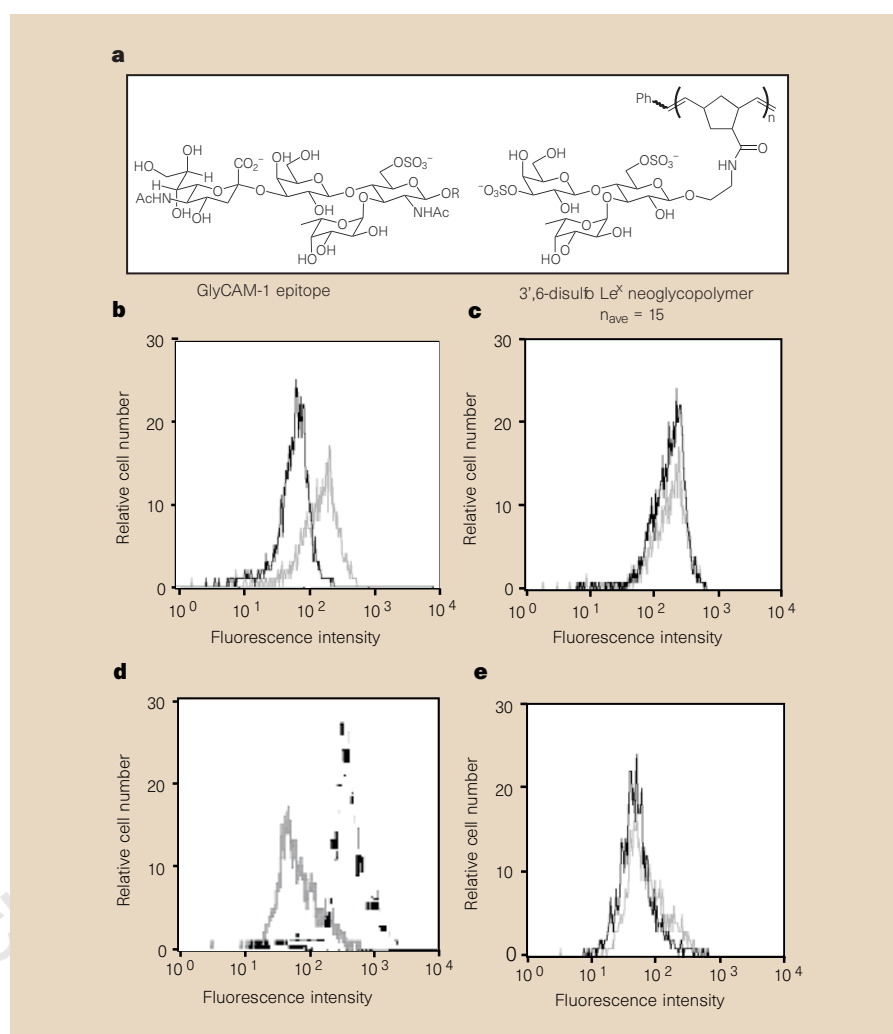


Figure 1 Effects of neoglycopolymer on human neutrophils. **a**, Chemical structures of an L-selectin binding epitope on GlyCAM-1 (Ac = acetyl) and the 3',6-disulpho Lex neoglycopolymer. **b**, Histogram overlay of L-selectin expression by untreated cells (grey) and cells treated with the 3',6-disulpho Lex neoglycopolymer (267 μM or 4 mM on a per saccharide basis; black). **c**, Histogram overlay of L-selectin expression by untreated cells (grey) and cells treated with the monomeric ligand 3',6-disulpho Lex(Glc)-β-OPr (4 mM) (black). **d**, Histogram overlay of Mac-1 expression by untreated cells (grey) and cells treated with 10 nM PMA (black). **e**, Histogram overlay of Mac-1 expression by untreated cells (grey) and cells treated with 3',6-disulpho Lex neoglycopolymer (267 μM or 4 mM on a per saccharide basis; black). Histogram depictions are representative of at least three experiments monitored by flow cytometry using a fluorescein isothiocyanate-conjugated anti-L-selectin antibody or a phycoerythrin-conjugated anti-Mac-1 antibody.

Such ligands might cluster L-selectin by mimicking either the mucins themselves or the multivalent display of mucin epitopes at the cell surface.

The L-selectin recognition element was based on the disulphated trisaccharide 3',6-disulpho Lewis x (Le^x), an analogue of the GlyCAM-1 capping group 6-sulpho sialyl Le^x4 (Fig. 1a). To assemble the recognition elements into a multivalent saccharide array⁶, we used ruthenium carbene-catalyzed ring-opening metathesis polymerization (ROMP)⁷, a technique that tolerates multiple unprotected functionalities and can produce polymers of consistent and controllable lengths. We produced synthetic ligands composed of about 15 monomer units, each substituted with the recognition epitope 3',6-disulpho Lex (Fig. 1a)⁸.

When we treated human neutrophils with these neoglycopolymers, L-selectin was lost from the cell surface in a dose-dependent manner (Fig. 1b). Soluble L-selectin was detected in the supernatant of treated cells; therefore the decrease on the cell surface was due to shedding.

In contrast, the monovalent ligand 3',6-disulpho Lex(Glc)-β-OPr did not cause L-selectin release (Fig. 1c). And a galactose-substituted oligomer, which does not bind L-selectin, had no effect.

Cellular activators, such as phorbol esters and chemotactic peptides, also induce the shedding of L-selectin; at the same time they dramatically increase the surface concentration of the β2-integrin Mac-1 (CD11b/CD18) (Fig. 1d), a protein involved in the next step of leukocyte

recruitment to the endothelium⁹.

To determine whether a similar mechanism was associated with neoglycopolymer-promoted L-selectin shedding, we monitored Mac-1 levels in neoglycopolymer-treated cells. Surprisingly, the neoglycopolymer had no effect on Mac-1 expression (Fig. 1e). Moreover, hydroxamic acid-based protease inhibitors, which diminish the activation-induced shedding of L-selectin¹⁰, do not prevent neoglycopolymer-induced shedding (data not shown).

The disparity in protease inhibitor activities raises the possibility that different proteases mediate the activation-dependent and independent events. These results indicate that molecules sharing aspects of physiological L-selectin ligands can induce L-selectin shedding, and that the release mechanism is distinct from that occurring following cellular stimulation.

Several unique features characterize neoglycopolymer-promoted shedding of L-selectin. The neoglycopolymers, like other selectin inhibitors, compete with natural ligands for selectin binding by non-covalent association. These agents can also facilitate a change in covalent bonding, thereby irreversibly altering the cell surface. Specifically, molecules that cause L-selectin release will diminish leukocyte rolling and the transendothelial migration that can follow. Concomitantly, they increase the concentration of soluble L-selectin that can act as an inhibitor of the rolling process. Moreover, when receptors are shed from the cell surface, the affinity of the soluble receptors for the multivalent ligand should decrease, allowing the multivalent ligand to function at substoichiometric levels.

Proteins that are shed from the cell surface form a diverse group. They include growth factors, cytokine receptors, cell adhesion molecules and leukocyte antigens¹. This diversity underscores the potential applications of receptor shedding in altering cellular responsiveness to specific ligands, or promoting responses at distal sites. We have demonstrated the possibility of devising molecules that selectively promote the shedding of L-selectin, suggesting a new strategy for generating anti-inflammatory agents. Because other proteins may undergo ligand-induced shedding, these results have broad implications for the investigation and manipulation of the cell surface and the extracellular environment.

Eva J. Gordon, William J. Sanders, Laura L. Kiessling

Departments of Chemistry and Biochemistry,
University of Wisconsin-Madison,
Madison, Wisconsin 53706, USA
e-mail: kiessling@chem.wisc.edu

3. Springer, T. A. *Cell* 76, 301–314 (1994).
4. Hemmerich, S., Leffler, H. & Rosen, S. D. *J. Biol. Chem.* 270, 12035–12047 (1995).
5. Walcheck, B. et al. *Nature* 380, 720–723 (1996).
6. Lee, Y. C. & Lee, R. T. *Acc. Chem. Res.* 28, 321–327 (1995).
7. Lynn, D. M., Kanaoka, S. & Grubbs, R. H. *J. Am. Chem. Soc.* 118, 784–790 (1996).
8. Sanders, W. J., Katsumoto, T. R., Bertozzi, C. R., Rosen, S. D. & Kiessling, L. L. *Biochemistry* 35, 14862–14867 (1996).
9. Kishimoto, T. K., Jutila, M. A., Berg, E. L. & Butcher, E. C. *Science* 245, 1238–1241 (1989).
10. Feehan, C. et al. *J. Biol. Chem.* 271, 7019–7024 (1996).

Thyroid cancer risk to children calculated

The Chernobyl reactor accident was followed by a sharp increase in the incidence of thyroid cancer among children and adolescents in Belarus (Belorussia) and Ukraine^{1,2}. Exposure to iodine-131 (¹³¹I) was responsible for most of the doses that affected the thyroids of these children; however, among evacuees, up to 40% of each dose could derive from other incorporated radionuclides and external exposures³. From the data set compiled after this incident, we estimated the increased risk of developing thyroid cancer after exposure to radioactive iodine. The figure we obtained for most of the affected regions fell within the 95% confidence interval of a previous follow-up of thyroid cancer after external exposures.

In Ukraine, more than 150,000 measurements of ¹³¹I activity in the thyroid were performed from mid-May to mid-June 1986. Collimators shielded against radiation from other parts of the body and from the environment during the taking of these measurements. Background signals were continuously monitored and subtracted from the results.

In a re-evaluation of these measurements⁴, we reconstructed average thyroid doses due to ¹³¹I exposure for three regions close to Chernobyl — Kiev, Zhytomyr and Chernigov, containing 4,406 settlements (villages or towns) — and for evacuees

(Table 1). Dose distributions in settlements with fewer than 12 thyroid measurements were estimated from settlements contributing many measurements by taking into account contamination by caesium-137 (¹³⁷Cs) of the ground, and the settlement's location relative to Chernobyl³.

For children less than three years old living in contaminated areas, the dose received exceeded that to adults by a factor of five, while among evacuees the child dose was about 20 times greater than that received by adults. Individual doses show large variability, but this contributes only slightly to the uncertainty of average thyroid dose estimates because of the size of this study. We estimate that a factor of 4 separates the 95% confidence boundaries of average thyroid doses.

For Belarus, deriving ¹³¹I activity concentrations in thyroids is more difficult because here the measurements of ¹³¹I activities in the thyroid were performed with non-spectrometric and uncollimated devices. However, ¹³¹I activity measurements in milk and soil are available^{5,6}. Thyroid doses were assessed for 812 settlements. In Bryansk, the most contaminated region of Russia, 14,000 measurements of ¹³¹I activities in thyroids were performed. Count rates detected with non-spectrometric devices were corrected for the contribution of radio-caesium in the body. Thyroid doses in 602 settlements were calculated by taking into account data on milk and soil contamination⁷.

Of 264 checked cases from Belarus and Ukraine, western European pathologists approved the thyroid cancer classification in 97% (ref. 8). In southern Ukraine, which we used as a control area, the thyroid cancer incidence among 9 to 18-year-olds increased by 70% from the period 1986–1988 to the period 1991–1995. This probably reflects both greater thyroid surveillance and Chernobyl radiation. In principle there might be a higher screening effect in the contaminated regions. However, the high frequency of metastases (these were found in lymph nodes in 66% of the children operated on in

Table 1 **Thyroid cancer risk among children born between 1971 and 1986**

Area	Children (× 10 ³)	Average thyroid dose (Gy)	Thyroid cancer cases 1991–1995	Observed/ expected ^a cases	Excess absolute risk (per 10 ⁴ person-year Gy)
Ukraine					
Zhytomyr oblast ^b	340	0.13	28	4	0.9
Kiev oblast	399	0.18	47	6	1.1
Chernigov oblast	273	0.09	33	6	2.2
Kiev city	581	0.05	67	6	3.8
30 km zone (evacuees)	20	0.92	12	30	1.3
Belarus					
Gomel/Mogilev oblasts ^c	76	0.73	89	56	3.2
Minsk city	357	0.08	41	6	2.3
Gomel city	113	0.40	72	30	3.1
Russia					
Bryansk ^c	169	0.12	31	9	2.7

a: Expected cases are calculated on the basis of the incidence in southern Ukraine of 4.2 cases per 10⁶ person-years

b: An oblast is an administrative region

c: Subarea of the region only

1. Hooper, N. M., Karran, E. H. & Turner, A. J. *Biochem J.* 321, 265–279 (1997).

2. Kansas, G. S. *Blood* 88, 3259–3287 (1996).

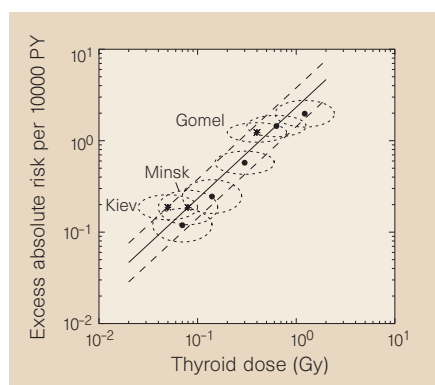


Figure 1 Excess thyroid cancer risk in the period 1991–1995 among people born between 1971 and 1986. The dots indicate average values of settlements in the categories of dose 0.05–0.1, 0.1–0.2, 0.2–0.5, 0.5–1.0, and 1–2 Gy; the stars show results for cities with large collective doses; PY = person-year. The solid line is the best estimate of the excess absolute risk per unit dose. Broken and dotted lines indicate 95% confidence ranges.

Kiev) indicates that most of the cases would have become obvious independent of any screening.

In southern Ukraine, the thyroid cancer incidence rate in 1991 to 1995 among the birth cohort 1971 to 1986 was 4.2 per 10^6 person-years⁹. Confidence intervals were derived from incidences in the period 1986 to 1988. In the northern part of Ukraine, the incidence differed by a factor of 1.5, and in Belarus and Bryansk by a factor of two from the incidence in southern Ukraine. The excess absolute risk shows no statistically significant deviation from a linear dose–response relationship (Fig. 1). A χ^2 -fit to data pooled for each of the three countries resulted in an excess absolute risk per unit dose of 2.3 (95% confidence interval: 1.4–3.8) per 10^4 person-year Gy (grays: a unit of absorbed dose). This is a factor of two lower than the best estimate derived from a previous study of thyroid cancer after external exposures where the period of follow-up ranged from five years to several decades¹⁰. However, the result lies within the 95% confidence interval of that study. During the observation period of our study, the excess thyroid cancer incidence was rising. So the excess absolute risk per unit dose is expected to increase for longer observation periods.

The excess relative risk per unit dose ranges between 22 Gy⁻¹ (Zhytomyr oblast) and 90 Gy⁻¹ in the study area. These results are higher than those obtained from longer observation periods after external exposures¹⁰, possibly because the population we studied is still young and therefore has very low baseline rates.

P. Jacob, G. Goulko, W. F. Heidenreich
GSF–National Research Center for Environment and Health, D-85764 Neuherberg, Germany
e-mail: jacob@gsf.de

I. Likhtarev, I. Kairo

Radiation Protection Institute, 252050 Kiev, Ukraine

N. D. Tronko, T. I. Bogdanova

Ukrainian Research Institute of Endocrinology and Metabolism, 254114 Kiev, Ukraine

J. Kenigsberg, E. Buglova, V. Drozdovitch, A. Golovneva

Scientific Research and Clinical Institute of Radiation Medicine and Endocrinology, 220600 Minsk, Belarus

E.P. Demidchik

Thyroid Cancer Center, 220600 Minsk, Belarus

M. Balonov, I. Zvonova

Research and Technical Center Protection, 197101 St Petersburg, Russia

V. Beral

Imperial Cancer Research Fund, Oxford OX2 6HE, UK

1. Kazakov, V. S., Demidchik, E. P. & Astakova, L. N. *Nature* 359, 21 (1992).
2. Likhtarev, I. A. *et al. Nature* 375, 365 (1995).
3. Goulko, G. M. *et al. Radiat. Environ. Biophys.* 36, 261–273 (1998).
4. Likhtarev, I. A. *et al. Health Phys.* 69, 6–15 (1995).
5. Drozdovitch, V. V. *et al. Radiat. Environ. Biophys.* 36, 17–23 (1997).
6. Gavrilin, Y. *et al. in The Radiological Consequences of the Chernobyl Accident*, 1011–1020 (EUR 16544, European Commission, 1996).
7. Zvonova, I. A. & Balonov, M. I. *in The Chernobyl Papers*, 71–126 (Research Enterprises Publishing Segment, Richland, Washington, 1993).
8. Williams, E. D. *et al. in The Radiological Consequences of the Chernobyl Accident* 699–714 (EUR 16544, European Commission, 1996).
9. Sobolev, B., Heidenreich, W. F., Kairo, I., Jacob, P., Goulko, G. & Likhtarev, I. *Radiat. Environ. Biophys.* 36, 195–199 (1997).
10. Ron, E. *et al. Radiat. Res.* 141, 259–277 (1995).

Sulphur isotope data consistency improved

The standard used for reporting relative sulphur isotope-abundance data has, historically, been troilite (FeS) derived from the Cañon Diablo meteorite, CDT. However, the isotopic inhomogeneity of this material — the variability in its sulphur-34/sulphur-32 isotope ratio is at least 0.4‰ — greatly exceeds the achievable analytical uncertainty of 0.05‰ (ref. 1). Accordingly, the International Union of Pure and Applied Chemistry (IUPAC) has recommended the setting up of a VCDT scale, which is expected to improve agreement between laboratories in measuring this ratio, $\delta^{34}\text{S}$.

Table 1 Examples of reference material

Name	NIST reference no.	$\delta^{34}\text{S}$
IAEA-S-1 silver sulphide	RM 8554	–0.3‰
IAEA-S-2 silver sulphide	RM 8555	~22‰
IAEA-S-3 silver sulphide		~32‰
NBS-123 sphalerite	RM 8556	~17‰
NBS-127 barium sulphate	RM 8557	~20‰
Soufre de Lacq elemental S	RM 8553	~16‰

A proposal that a VCDT scale be established was originally made in 1993, at the Consultants' Meeting on Stable Isotope Standards and Intercomparison Materials, in Vienna, sponsored by the International Atomic Energy Agency (IAEA)². The scale is defined by assigning a $\delta^{34}\text{S}$ value of –0.3‰ exactly (relative to VCDT) to the internationally distributed silver sulphide reference material IAEA-S-1 (formerly called NZ1). This decision was based upon measurements of IAEA-S-1 in several laboratories using SO_2 ($\delta^{34}\text{S} = -0.25 \pm 0.31\text{‰}(2\sigma)$ for 17 laboratories³) and SF_6 ($\delta^{34}\text{S} = -0.3 \pm 0.1\text{‰}(2\sigma)$ for three laboratories³).

This recommendation by the IAEA consultants found favour with the Commission on Atomic Weights and Isotopic Abundances of IUPAC at a meeting in Guildford in the United Kingdom, in August 1995 (ref. 4). Accordingly, IUPAC recommended that:

- the use of meteoritic troilite and the reporting of $\delta^{34}\text{S}$ data relative to CDT be discontinued;
- relative sulphur isotopic abundance data be reported relative to VCDT; and
- the VCDT scale be established in laboratories through the use of IAEA-S-1 silver sulphide.

In addition, authors are encouraged to report $\delta^{34}\text{S}$ values for internationally distributed sulphur isotopic reference materials (Table 1) if they have been analysed. Sources of isotopic reference materials include:

- the National Institute of Standards and Technology, Standard Reference Materials Program, Room 204, Building 202, Gaithersburg, Maryland 20899-0001, United States. Tel: +1 301 975 6776; fax: +1 301 948 3730; e-mail: SRMINFO@enh.nist.gov and

- the International Atomic Energy Agency, Section of Isotope Hydrology, Wagramerstrasse 5, PO Box 100, A-1400 Vienna, Austria. Tel: +43 1 206021740; fax: +43 1 20607; e-mail: RIALIHL@iaea.org

Tyler B. Coplen

US Geological Survey, 431 National Center, Reston, Virginia 20192, USA
e-mail: tbcoplen@usgs.gov

H. Roy Krouse

University of Calgary,
Department of Physics and Astronomy, Calgary, Alberta T2N 1N4, Canada
krouse@acs.ucalgary.ca

1. Beaudoin, G., Taylor, B. E., Rumble III, D., & Thiemens, M. *Geochim. Cosmochim. Acta* 58, 4253–4255 (1994).
2. *Reference and Intercomparison Materials for Stable Isotopes of Light Elements* (International Atomic Energy Agency, Vienna, TECDOC-825) 7–12 (1995).
3. Robinson, B. W. *in Reference and Intercomparison Materials for Stable Isotopes of Light Elements* (International Atomic Energy Agency, Vienna, TECDOC-825) 39–45 (1995).
4. Krouse, H. R., & Coplen, T. B. *Pure Appl. Chem.* 69, 293–295 (1997).

Rise and fall of an academic empire

New Atlantis Revisited: Akademgorodok, the Siberian City of Science

by Paul R. Josephson

Princeton University Press: 1997. Pp. 351.

\$39.50, £27.50

David Joravsky

This history opens with Francis Bacon as its prophet, the “father of the modern inductive method of reasoning” and visionary of “the new Atlantis”, a city of learned men that endlessly expands knowledge, thereby “enlarging the bounds of human empire, to the effecting of all things possible”. Bacon’s “modern inductive method” is never spelled out, so one can only guess whether it might have clarified the blooming, buzzing confusion that a sceptic finds in Soviet history.

But his vision of the ‘new Atlantis’ is described, and permeates the history as pure vision, admirable though ‘utopian’, spoiled in realization by social and political circumstances. Paul R. Josephson cries alarm when Bacon’s ‘human empire’ reaches Lake Baikal, bringing irremediable ruin along with paper mills, and when such empire grasps at the rivers of Siberia, seeking to reverse their course, to ease the catastrophic overload that farm and factory and swelling city have placed on the precarious water supply of Russia and central Asia. But he never stops to ponder the linkage of power and knowledge in the vision itself, either in Bacon’s monarchical version or in that of Soviet communists such as M. A. Lavrentev, mathematician and academic empire builder. It was Lavrentev who organized one of the last great crash programmes of Soviet power: Akademgorodok, the ‘new Atlantis’ in Siberia, a city of science that flattened virgin forest.

Lavrentev’s grand project of Akademgorodok was part of Khrushchev’s effort to rescue the Communist Party and the country from the ruin that threatened after Stalin had pressed toward self-destructive limits in his own grand projects; never mind the hugely destructive side effects. Josephson highlights the partnership of Lavrentev and Khrushchev, empire-building scientist and supreme political boss, but he persistently makes ‘science’ and ‘the Party’ seem distinct and separable. Lavrentev joined the party under Stalin, a crucial fact I learned from a Soviet biography. It is not mentioned, and therefore not interpreted by Josephson. In his account, gigantomania and *shтурмовщина* (“taking by storm”, or the sporadic rushing to achieve goals by enthusiastic assault) were faults not of ‘science’ but of ‘the Party’ or ‘the Soviet system’.

He evades the obvious question that hangs over his scientist heroes, who threw themselves into a Soviet-style project to “overtake



Akademgorodok: one of the last great crash programmes of Soviet power.

and surpass” Western science. Were they not Soviet-style heroes, committed to ‘great leaps forward’ along with their political chiefs, to flights of endeavour that would fall very hard because they aspired so high, so fast, with ‘blue-sky’ numbers for goals and costs?

“Soviet-style infrastructure”, writes Josephson, “sabotaged utopian scientific visions.” He ignores the possibility that these were two aspects of one process, a mutually reinforcing spiral of crippling realities and escapist revolts. Trying to be fair, he promises to make judgements “from the perspective of Soviet standards”, but he is irresistibly drawn to comparison with world standards. “Overtake and surpass” was after all a central goal that Soviet communists set for themselves.

He is happy to tell how nuclear physicists in Akademgorodok sometimes surpassed Western pacesetters in their field, and so unavoidably he must tell how they subsequently fell behind. But he does not ask whether this trajectory calls into question the wisdom of the initial plunge into Siberia. Indeed, such trajectories of great achievement followed by greater disappointment are a feature of *shтурмовщина*, and may be a pattern of Soviet history in general.

Gorbachev’s ‘reconstruction’ that turned into collapse is a spectacular case in point.

The records of aspiration and achievement that Josephson reports for biology, computer science, economics and sociology in the ‘new Atlantis’ raise more interesting questions than its physics, for they were more intricately entangled with social practice at large, each visibly shaping the other. Economics provides a particularly striking instance. Josephson retells the story of linear programming, invented by the mathematician Kantorovich for the optimum use of eight machines processing five types of wood. When economists tell the story, they show that Kantorovich was also reinventing the concept of marginal utility — as Stalin, let me add, reinvented marginal calculation in land use after he had denied the existence of land rent in a socialist system. Calculation of efficiency is a fiendish problem. It can be at odds with considerations of social justice after the revolution as well as before, especially in ‘developing’ countries. And interests of power can raise hell all over.

For nearly 70 years, Soviet economists and bosses played intricate games among themselves, avoiding frank acknowledgement of a

basic question: how could they extend the concept of marginal utility to the calculation of costs and benefits throughout the system without abandoning worthy public projects — such as basic research — which are often driven by intuitions and enthusiasms that defy calculation of marginal utility? And how can truly worthy public projects be distinguished from interests of power, or from ‘pork’ as Americans call appropriations by particular coteries of patrons and clients?

As Josephson tells his stories, the painful complexity of these questions emerges less in the glory years of the reckless leap to Atlantis than in the aftermath, especially in the recent surrender to ‘the market’, a euphemism in post-communist Russia for enormously bloated pork-barrel politics. He points out that the market has even undermined intellectual freedom by the near destruction of state funding for science, obliging researchers to scramble for the support of profit-seeking sponsors.

Josephson evokes the excitement of the heaven-storming years in the Siberian science city, before Khrushchev’s comrades cast him aside to enter the ‘period of stagnation’. Much of the book declines with its protagonists into obsessive concern for networks of patrons and clients and intricate intrigues, with a deadening effect not only on the vitality of the system but also on the reader’s interest. The challenge of such personal networks is to discover in their historical development patterns of institutional change that transformed elements of a ruling party into elements of a ruling class, while changing hard-driving scientist entrepreneurs of Lavrentev’s type into grey-faced survivors, profit-seeking hustlers or émigrés.

I construe this oversimplified phylogeny from torrents of anecdotal evidence in Josephson’s book. Too often he seems unaware of a need to seek patterns of institutional evolution, whether through Baconian induction or wilful construction. He writes as if the institutions in question were, on one side, multitudinous bureaucratic structures such as the Academy of Sciences, its branches and institutes, and, on the other side, a single entity, ‘the Party’. Soviet rhetoric commonly invoked the bizarre image of a huge, complex society dominated by a monolith, and many people still talk that way, now blaming rather than revering ‘the Party’, but still avoiding acknowledgement that it too was a congeries in the process of change along with the organizations that were supposed to be controlled by ‘the Party’.

The most amusing instance of such talk comes in an interview with the one-time Party boss of Akademgorodok. Josephson gives a vivid picture of a smooth, effective operator, who served his local constituents well until 1965, when the central bosses in the Politburo kicked out Khrushchev, downgraded the project of a ‘new Atlantis’ and abolished the

special post of the local boss in Akademgorodok. “The Party”, he told Josephson, “didn’t understand. To them science was just one aspect of various slogans.” That reference to ‘the Party’ as an alien entity, ignoring the speaker’s active membership in it, is not only comic but suggestive, inviting the scholar to start picking apart a heterogeneous institution that aspired to be a monolith.

The scientific front, as the comrades used to say, is a good place to seek the complex interactions that moved Soviet institutions through revolutionary spasms to stagnation, and thus to terminal spasms and collapse, engendering crony capitalism in the ruins of crony socialism. □

David Joravsky is in the Department of History, Northwestern University, Evanston, Illinois 60208-2200, USA.

Pairs of genes

Twins

by Lawrence Wright

Weidenfeld & Nicolson/Wiley: 1997. Pp. 202.

£14.99, \$22.95

John Galloway

W. R. Inge, a former dean of St Paul’s Cathedral in London, once remarked that the distinction between literature and journalism had become blurred. This might now be said about science and journalism. Scientists are increasingly taking up journalism, and here a journalist, Lawrence Wright, has taken to science.

His book is about the relative extent to which people’s lives and personalities are influenced by their genes and surroundings. How much do genes pre-programme people, deciding their intelligence, success or

happiness? Are all our choices in some way pre-determined? As the title suggests, the recurring theme of the book is how scientists use twins to disentangle these influences. But they may not be as good a model as scientists assume. Research using twins relies on knowing their zygosity, and errors of up to 30 per cent may occur in using chorionic membranes to determine this.

How good a book is *Twins*? As a measured analysis of a complex and controversial field, it is not very good. Journalists are concerned mainly with putting readers in touch with issues, and as a result they tend to report other people’s views rather than develop their own. The book is essentially a collection of the sometimes divergent views of the field’s main players about the significance of genes. Several reviewers have failed to make this point, for which the publisher’s blurb has to take some responsibility. It states, apparently as demonstrable fact, that “intelligence, political orientation, careers, suicide, are largely under genetic influence”, whereas in reality the book shows a lack of consensus among those most closely involved.

The notion of identical twins separated at birth and raised in very different families is seductive to scientists, although disturbing to the rest of us. Interestingly, it has its origin in literature. Mark Twain’s *The Prince and the Pauper* and Pudd’nhead Wilson are variations on this theme. In the latter, the newborn sons of a slave and his master are swapped at birth, allowing an exploration of the interplay between heredity and social circumstance.

Science caught up with art when the British psychologist Cyril Burt claimed to have traced more than 40 pairs of identical twins that had been separated at birth and brought up apart. Wright relates that, using psychological testing, Burt reckoned the split



Models of nature: identical twins are seductive scientific subjects.

HUTTON GETTY

in influence between genes and social environment was about 70:30. But doubt has been cast on this research: it has been suggested that some of the data were “made up” to buttress Burt’s view that genes encode the social order.

Peter Neubauer left less to chance in his study. Working through the Louise Wise Adoption Agency in New York, he carried out a long-term social and psychological experiment. He arranged for twins — and on one occasion triplets — to be separated and adopted by families from different social backgrounds. Neither the families nor the children were told. The ‘experiment’ came to light when two of the triplets accidentally discovered each other, and then found the third, a story suspiciously like Ken Follett’s novel *The Third Twin*.

Thomas Bouchard of the University of Minnesota made his reputation by studying the similarities and differences of separated twins whom he reunited. Extreme cases of similarity exist in which the twins’ lives seem to match in uncanny detail, even down to the clothes they wore when they first met as adults.

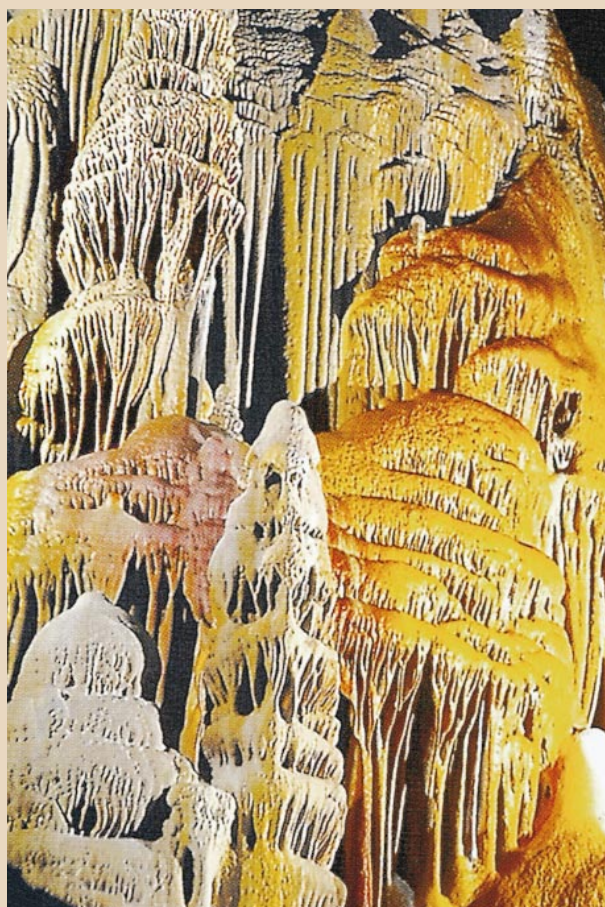
Although these findings appear astonishing at first sight and make great newspaper stories, what do they really mean? How many are just coincidences? Some are probably secondary to physical characteristics, which certainly are genetic in origin. Should similar research not be carried out using people less closely related? If scientists believe that genes do predispose people to wearing particular clothes, they should use techniques similar to those used to identify the genes suspected of predisposing towards inherited conditions such as asthma. According to Wright, Bouchard thinks it is too much effort to look at such control populations: “What could it tell you?”

What do scientists mean by environment? A lot of research concentrates on the differences between the family and the wider world. But environment before birth may be decisive, a point made forcefully by Bernard Devlin *et al.* (*Nature* **388**, 468; 1997). These findings were based on a meta-analysis of more than 200 studies of the heritability of IQ. As IQ and its heredity feature strongly in *Twins*, Wright should have said more about its provenance as a psychological and social indicator and the robustness of its measurement. And, if all or part of it can be inherited, what exactly are the genes coding for?

One plausible suggestion is that they control the speed at which the brain processes information. Wright does not say just how great the genetic differences are between people. In fact the differences appear to be rather subtle: about one base in a thousand in the three billion bases in the human genome, but with a tendency for differences to be concentrated in non-coding DNA.

Psychology ought to be the science of individuals: it should examine what makes people

Underground revelations



The Fiery Temple of the Cave God, a colourful ‘flowstone’ cascade in Three Fingers Cave, Guadalupe Mountains, New Mexico, is one of more than 250 mineral-deposit formations that see photographic light of day in the second edition of *Cave Minerals of the World* (National Speleological Society, \$70) by Carol Hill and Paolo Forti. The book describes in detail the wide range of mineral types found in caves, and provides a guide to the classification of secondary mineral deposits, or ‘speleothems’. The authors then stick their necks out and draw up a list of what they consider to be the top ten caves in the world in terms of mineral deposits.

different, even unique. Too often twins are used as convenient models for studying the human condition in its generality. Perhaps psychologists might study twins for themselves, and look at their developing personalities, their mutual social interaction and the social dynamics both inside and outside their families. It would probably be better for twins; it might even be better science.

The book does not quite work either as journalism or as a critical survey. What characterizes good journalism is that it makes its point on a first reading by reducing and simplifying the information content to the minimum. There is too much information here for that, but too little to give a clear perspective of the field. The science Wright describes is at best tentative and provisional, a point that may be lost, sometimes deliberately, on society in general and its leaders in particular. And this is a field where, as Wright argues persuasively, science has a tendency to lead to social theory and sometimes politics only too readily. It is worth remembering the words Bertolt Brecht put into the mouth of Galileo: “The aim of science is not to open a door to infinite wisdom but to set a limit to infinite error.” □

John Galloway is in the Team Development Unit, Eastman Dental Hospital, 256 Gray’s Inn Road, London WC1X 8LD, UK.

True stories

A Novel Defense of Scientific Realism

by Jarrett Leplin

Oxford University Press: 1997. Pp. 204.
\$39.95, £35

Stathis Psillos

Modern science has transformed the way we think of the world. Nature is no longer taken to be as our senses tell us it is. Entities and mechanisms invisible to the naked eye, such as electromagnetic waves, electrons, protons, neutrinos and DNA molecules, to mention but a few, are said to populate the world and cause the phenomena we observe.

But why should we take scientific theories to be true, or nearly so? Why can we not just take them to be mere instruments for the systematization and prediction of observable phenomena, without attributing reality to the invisible entities they posit? Or could we not just suspend our judgement as to the truth of the assertions the theory makes about invisible entities, and believe only that the theory is empirically adequate, that is, that whatever it says about the observable phenomena — and

only this — is true?

Scientific realists argue that mature and genuinely successful scientific theories should be accepted as nearly true. The main defence of this optimistic attitude is known as 'the no miracle argument' because it is based on Hilary Putnam's slogan that "scientific realism is the only philosophy of science that doesn't make the success of science a miracle".

Modern defenders of scientific realism, most notably the philosopher Richard Boyd at Cornell University, have based their defence on the idea that the impressive predictive and explanatory successes of scientific theories would remain unaccounted for unless we accept that the entities, processes and causal mechanisms they posit to operate behind the phenomena are real. They dismiss instrumentalist accounts of scientific theories by pointing out that they leave the success of science unexplained. If theories are merely 'black boxes', the only virtue of which is that they offer the most economical classification of the observable phenomena, then there is no

reason to expect that they can be, as the French philosopher and scientist Pierre Duhem put it, "prophets for us". To counter that these black boxes are empirically adequate — that they save all phenomena — would not be much of an improvement on the instrumentalist position. For what needs explanation is precisely the fact, if it is a fact, that scientific theories save the phenomena. To say that they do is merely to assert what needs to be explained.

The realist explanation of the success of science does not go unchallenged. One line of criticism is that it is too easy to obtain empirical success: just 'write into' the theory the right observational consequences. Then the theory would not fail to predict them. But realists are quick to regiment their argument. There is a kind of prediction that can support only a realist understanding of the theory that entails it: the prediction of novel phenomena. For there is no explanation of why the theory predicts the existence of a novel phenomenon, other than to say that the phenomenon is brought about by

the theoretical mechanisms posited by the theory. A novel prediction is typically taken to be the prediction of a phenomenon whose existence is ascertained only after a theory suggests its existence.

But this cannot be the whole story because theories also receive support from explaining phenomena that are already known. So some philosophers argue that the 'temporal view' of novelty is inadequate and should be replaced by a 'novelty in use' view: a prediction of an already known phenomenon can be 'use novel' with respect to a theory provided that information about this phenomenon was not used in the construction of the theory. Yet it has been notoriously difficult to make precise the intuitive idea of 'use novelty'.

A Novel Defence of Scientific Realism takes up this challenge and meets it beautifully. Jarrett Leplin analyses 'novelty' by reference to two requirements: independence and uniqueness. The core idea is that a prediction of a phenomenon O, be it already known or hitherto unforeseen, is novel for a theory T if no information about O is necessary for the prediction of O by T and if there is no other theory available that explains why O should be expected.

Leplin puts his analysis of novelty to work in his defence of scientific realism. His argument is not new. It is a variant of the known line that a realist understanding of theories that entail novel predictions offers the only explanation of their capacity to yield such predictions. But I think this line is right, and here Leplin does a thorough job in developing the realist position and blocking well-known counter-arguments.

However, the realism he wants to entertain is rather weak. His 'minimal epistemic realism' is committed merely to the claim that "there are possible empirical conditions that would warrant attributing some measure of truth to theories — not merely to their observable consequences, but to theories themselves". As he knows, many realists would aspire to more. They would, for instance, try to defend the more substantive thesis that the distinctive mode of inference involved in the generation and acceptance of explanatory hypotheses in science, known as inference to the best explanation, is conducive to truth. Here Leplin parts company with his fellow realists and wishes them *bon voyage*. Their voyage has to be successful if the rationality of belief in unobservable entities is to be defended. Consequently, more work needs to be done to defend a full-blooded realism. But this book is a splendid starting point. □

Stathis Psillos is in the Department of Philosophy, London School of Economics, Houghton Street, London WC2A 2AE, UK.

At a glance

Sex, Color, and Mate Choice in Guppies

by Anne E. Houde

Princeton University Press: 1997. Pp. 210. \$49.50, £35 (hbk); \$19.95, £14.95 (pbk)

It is hard not to like a book with an appendix titled "How to build a better bordello", although the subjects on display here are male guppies. Building an effective bordello, or at least a reliable way to determine mating preferences, is important to biologists who study guppies, and, like much of the rest of this book, turns out to be of interest to researchers into sexual selection in other species as well.

For studies of adaptation in the wild in general and of mate choice in particular, guppies have been a model system for many years, and the author synthesizes the results of decades of research. She carefully distinguishes between speculation and fact, and provides one of the most lucid analyses of current mate-choice models I have read. Literature cited on sexual selection is impressively broad for vertebrates, less so for insects and other invertebrates.

Is the book just for 'guppy people'? Yes and no; researchers anxious to keep up with the nuances of the latest experiments on the significance of the amount of orange area relative to the intensity of orange will find details of limited interest to the rest of us. The last few chapters, however, particularly the concluding section, provide an inspiring guide to future directions in behavioural ecology.

Marlene Zuk Department of Biology, University of California, Riverside, California 92521, USA.

The Handicap Principle: A Missing Piece of Darwin's Puzzle

by Amotz Zahavi and Avishag Zahavi

Oxford University Press: 1997. Pp. 286. \$30, £18.99

At first sight, the peacocks tail emblazoned across the dustjacket is an affront to evolution by natural selection. Surely such a cumbersome trait could not arise by 'survival of the fittest'? Such naive criticisms were pre-empted by Darwin with his theory of sexual selection.

In the mid-1970s, Amotz Zahavi published a controversial series of papers in which he claimed that all previous theories of sexual selection were flawed. Instead, he proposed his handicap principle. Moreover, he asserted that Darwin's demarcation between natural and sexual selection was incorrect, the true distinction being between natural and 'signal' selection.

This insightful yet accessible book explains the handicap principle and applies it to the analysis of a wide variety of behavioural phenomena: altruism, mate choice, the signalling between predator and prey, to name but a few. Many of the Zahavis' ideas are inspired and should be required reading for all zoologists and interested lay people. Unfortunately, this high praise must come with at least one caveat: the book is extremely tendentious. Although the authors provide a rigorous defence of their own position, rival theories are often unfairly dismissed with glib and specious arguments.

Overall, this is a beguiling tale, pleasantly illustrated, which is of genuine and general importance to our understanding of evolution and animal behaviour.

Steven Siller British Telecommunications Laboratories, Martlesham Heath IP5 3RE, UK.

The hydrogen hypothesis for the first eukaryote

William Martin* & Miklós Müller†

* Institut für Genetik, Technische Universität Braunschweig, Spielmannstrasse 7, D-38023 Braunschweig, Germany

† The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

A new hypothesis for the origin of eukaryotic cells is proposed, based on the comparative biochemistry of energy metabolism. Eukaryotes are suggested to have arisen through symbiotic association of an anaerobic, strictly hydrogen-dependent, strictly autotrophic archaeobacterium (the host) with a eubacterium (the symbiont) that was able to respire, but generated molecular hydrogen as a waste product of anaerobic heterotrophic metabolism. The host's dependence upon molecular hydrogen produced by the symbiont is put forward as the selective principle that forged the common ancestor of eukaryotic cells.

Unicellular eukaryotes (protists) that possess neither mitochondria nor hydrogenosomes—the double-membrane-bounded, H_2 - and ATP-producing organelles of amitochondriate protists¹—have figured prominently in hypotheses for eukaryotic origins^{2–4}. Candidates for the most primitive contemporary eukaryotes have been sought among these groups, because they are eukaryotes but are devoid of organelles that descend from free-living eubacteria under the endosymbiont hypothesis⁵. Two main hypotheses currently explain how such a hypothetically ancestral, organelle-lacking eukaryote might have arisen. The Archezoa hypothesis^{6,7} is founded in comparative cytology. It posits that eukaryotes and archaeobacteria share a common ancestor, and that the ancestral eukaryote (an archezoon) arose directly from that stem by evolving a nucleus, a primitive cytoskeleton, and endocytosis. A descendant of that archezoon is suggested to have endocytosed a eubacterium that became the mitochondrion, while others remained amitochondriate. The alternative ‘fusion’ hypothesis draws primarily upon molecular phylogenetic data⁸. It accepts Archezoa as the starting point of eukaryotic evolution, but derives them from a fusion event between an archaeobacterium and a eubacterium. In some formulations, fusion involved engulfment of the archaeobacterium by the eubacterium, in other words an endosymbiotic origin of the nucleus^{9,10}. The resulting chimaera is proposed to have then evolved eukaryotic structures and acquired the mitochondrion as above.

The Archezoa hypothesis offers plausible and explicit mechanisms for the origin of eukaryotic cellular features, yet cannot directly account for the findings (1) that many eukaryotes that lack both mitochondria and hydrogenosomes possess nuclear genes thought to be of eubacterial origin^{11–15} and (2) that among contemporary archaeobacteria, none can be found that possesses cytological structures that can be meaningfully homologized to those typical of eukaryotes. The fusion hypothesis (and variants thereof³) accounts directly for eubacterial genes in Archezoa, yet offers no plausible explanation for the biological context of interkingdom fusion and fails to offer explicit mechanisms for the origin of eukaryotic structures other than the endoplasmic reticulum and nucleus.

Both hypotheses embrace the view that the host of mitochondrial symbiosis was a eukaryote. Neither hypothesis examines specifically what type of energy metabolism the ancestral eukaryote and its antecedent(s) may have had, but rather assume that the host was heterotrophic before the acquisition of mitochondria^{4,5,16,17}. Here we summarize energy metabolism in non-photosynthetic eukaryotes and put forward an explicit inference as to its ancestral state. The result of that inference is a hypothetical, primitive eukaryotic cell with surprising attributes.

Eubacterial energy metabolism in eukaryotes

Eukaryotes that do not possess functional plastids are heterotrophic: they satisfy their ATP needs through the oxidative breakdown of reduced organic compounds (Fig. 1). Glycolysis (the Embden–Meyerhoff pathway) is the backbone of eukaryotic energy metabolism: one mol glucose is oxidized to pyruvate with the help of NAD^+ with a net yield of 2 mol ATP. In mitochondriate eukaryotes, pyruvate is usually further oxidized in the mitochondria through the pyruvate dehydrogenase complex (PDH), the Krebs cycle and O_2 respiration to yield CO_2 and water under the production of an additional 34–36 mol ATP per mol glucose. Amitochondriate eukaryotes meet their energy needs through anaerobic fermentations^{18–21}. They also obtain 2 mol ATP from glycolysis, but they differ from mitochondriate eukaryotes with respect to the fate of pyruvate. In amitochondriate eukaryotes, pyruvate is metabolized through pyruvate:ferredoxin oxidoreductase^{18–21} (PFO), rather than through PDH. In eukaryotes that lack organelles involved in core metabolism (type I amitochondriate eukaryotes^{18–21}), cytosolic PFO decarboxylates pyruvate, yielding reduced ferredoxin and acetyl-CoA. The latter is converted into a mixture of ethanol and acetate, the relative amounts of which depend upon environmental conditions, yielding between 0 and 2 additional mol ATP per mol glucose (Fig. 1a). In amitochondriate eukaryotes that harbour hydrogenosomes (type II amitochondriate eukaryotes^{18–21}), cytosolic pyruvate is imported into the organelle, where PFO converts it to CO_2 , acetyl-CoA and reduced ferredoxin. Ferredoxin is reoxidized by hydrogenase, producing the H_2 characteristic of the organelle. Per mol glucose, pyruvate metabolism in hydrogenosomes yields two additional mol ATP and two mol each of H_2 , CO_2 and acetate as waste products (Fig. 1b).

Whereas the endosymbiont hypothesis readily accounts for the eubacterial ancestry of mitochondrial energy metabolism⁵, the evolutionary origin of energy metabolism in amitochondriate protists has been more elusive. But, as summarized below, recent data suggest that it, too, is of eubacterial origin (in contrast to the archaeobacterial ancestry presumed for various components of the eukaryotic genetic apparatus^{3,14,22–25}). Because molecular data indicate that hydrogenosomes and mitochondria share a common ancestor^{26–29}, and because PFO and other enzymes of hydrogenosomes are of eubacterial ancestry^{19,30}, a case can be made for a eubacterial origin of (at least major segments of) energy metabolism in type II amitochondriate protists. This view is furthermore supported by the findings that some contemporary proteobacteria³¹ and cyanobacteria³² (1) can grow aerobically or anaerobically, (2) possess respiratory chains and hydrogenase and (3) possess

homologues of PDH and PFO, in turn suggesting that the common ancestor of hydrogenosomes and mitochondria did as well. Molecular data for enzymes of type I amitochondriate protists, such as PFO and bifunctional aldehyde/alcohol dehydrogenase in both *Giardia lamblia* and *Entamoeba histolytica* suggest that these enzymes are eubacterial, rather than archaeobacterial in origin¹⁵. Acetyl-CoA synthetase (ADP-forming) found in type I amitochondriate protists³³ is common among archaeobacteria³⁴, although some eubacteria are known that also possess this enzyme³³. Other nuclear genes of amitochondriate protists, in addition to those of PFO-related pathways, are thought to descend from eubacteria rather than from archaeobacteria^{11–14}. Notably, such genes include several enzymes of the glycolytic pathway from glucose to pyruvate in the eukaryotic cytosol^{12,13,19,35,36}.

At face value, these data suggest that (1) many, and probably all, groups of amitochondriate protists harboured eubacterial symbionts in their evolutionary past (2) that the enzymes essential to all three known types of eukaryotic energy metabolism were acquired from eubacteria and (3) that the free-living common ancestor of hydrogenosomes and mitochondria was capable of producing sufficient ATP both in anaerobic and aerobic environments. The simplest interpretation of these findings is that the three forms of energy metabolism found in eukaryotes today were inherited from the common ancestor of hydrogenosomes and mitochondria, which possessed the enzymes necessary to perform all three. From that it would follow that in the case of type II amitochondriate protists, the respiratory pathway and hydrogenosomal genome have been lost, whereas in the case of type I amitochondriate protists the entire organelle has additionally been lost. The phylogenetic distribution of type I and type II amitochondriate protists across ribosomal RNA phylogenies indicate that these losses have occurred many times in independent eukaryotic lineages^{3,7,19,21,23,37}. However, for the purposes of this paper, the order of these losses is irrelevant.

Metabolism in the context of symbiont origins

Traditional views on mitochondrial origins posit that their benefit to the host was increased efficiency of ATP production through respiratory carbohydrate breakdown. However, this generally accepted premise carries several tenuous corollary assumptions, most notably (1) that the host was unable to synthesize sufficient amounts of ATP by itself, (2) that the symbiont synthesized ATP in amounts exceeding its needs and (3) that the symbiont could export ATP to its environment, so that the host could realize this benefit. These phenomena are unknown among contemporary

cells, suggesting that ATP itself is unlikely as an initial symbiotic benefit. If not ATP, then what? Attempting to infer the context of benefit in an ancient symbiosis is necessarily speculative, but deserves exploration.

What might the symbiont have needed, what might have it been able to provide? From molecular phylogeny we can assume that it was a member of the α -proteobacteria^{5,14,26–29}, and that it therefore may have been photosynthetic or non-photosynthetic, autotrophic (able to satisfy its carbon needs from CO₂ alone) or heterotrophic, anaerobic or aerobic, or all of the above, as is the case for many contemporary representatives of the group, such as *Rhodospirillum rubrum*³⁸. From the previous section, we posit that the symbiont possessed (at least) PDH, a Krebs acid cycle, a complete respiratory chain and all the enzymes for energy metabolism as are found in amitochondriate protists. In order to grow, such a bacterium needs reduced organic compounds, but has little to offer the host other than waste products of its metabolism: CO₂ in the case of aerobic respiration, CO₂, H₂ and acetate in the case of 'hydrogenosomal' metabolism. Thus, if the context of symbiosis was metabolic, there are two possibilities: the host either could have (1) provided benefit to the symbiont in the form of reduced carbon substrate, or (2) reaped benefit from waste products of the symbiont's metabolism.

Alternative (1) is unlikely, because among the plethora of lithotrophic (generating ATP through redox reactions) and heterotrophic pathways known among contemporary archaeobacteria³⁴, only two produce reduced carbon compounds: heterotrophic fermentation and methanogenesis. Fermentation is unlikely as a benefit from host to symbiont, because if both grew heterotrophically, competition, not symbiosis would have ensued. Methane, by contrast, is the sole energy source of obligate methanotrophic α -proteobacteria³⁹. It is possible, but highly unlikely, that such was the initial context of host–symbiont association. This is because contemporary methanotrophy is strictly dependent upon molecular oxygen³⁹, whereas contemporary methanogens are strict anaerobes^{34,40}. An intimate cellular association of the type necessary to generate endosymbiosis cannot be construed, and has not been observed in natural communities⁴¹.

Alternative (2), however, unearths the many plausible benefits of hydrogen. Many archaeobacteria are strictly dependent upon H₂ for their ATP production^{34,40}. Moreover, for many methanogens (the strictly lithoautotrophic forms), H₂O and CO₂ are the sole source of both energy and carbon^{34,40}, whereas others can utilize alternative carbon sources such as methylamine, formic acid and acetate (all of which are waste products of eubacterial metabolism), and a few can grow on acetate alone^{37,40}. For methanogens, all three waste products

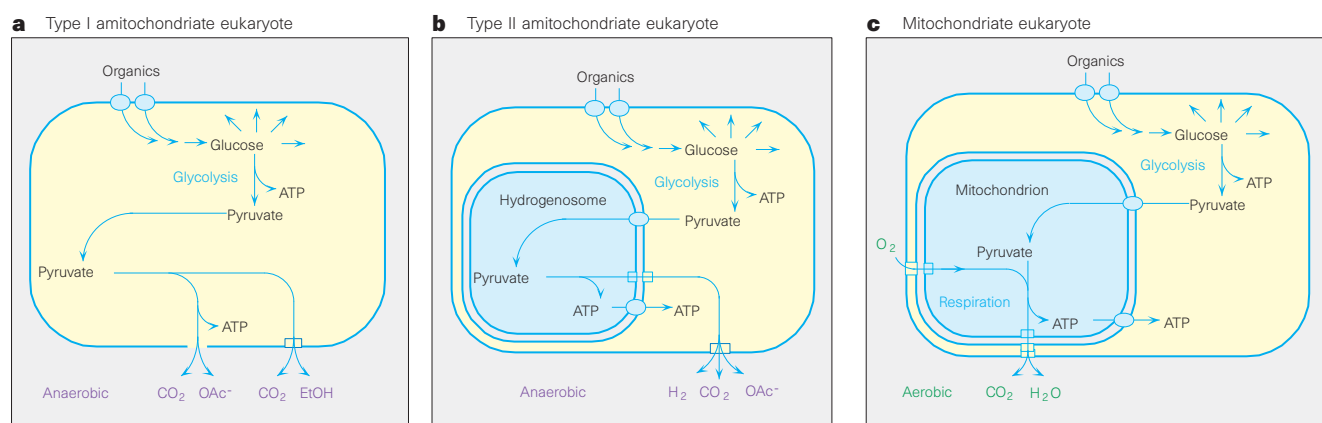


Figure 1 Schematic summary of forms of energy metabolism among heterotrophic eukaryotes (see refs 18 and 19 for details). OAc[−], acetate; EtOH, ethanol; ATP, adenosine triphosphate. Respiration in the figure designates the combination of oxidative decarboxylation by PDH, Krebs cycle (also known as citric acid

cycle, tricarboxylic acid cycle, TCA cycle) to produce NADH + H⁺ and FADH₂, and the respiratory electron transport chain that donates electrons and protons to O₂, yielding ATP through oxidative phosphorylation.

of the symbiont's anaerobic metabolism are fuel for life. Furthermore, methanogens are strictly anaerobic, congruent with the conditions under which both hydrogenosomes and free-living eubacteria produce H_2 . This is intriguing, but is it realistic? Are associations between methanogens and hydrogen-producing organisms observable today? Yes, abundantly so. Anaerobic syntrophy between methanogens and hydrogen-producing organisms has been known for many years⁴² and has been studied in some detail^{37,43–45}. It is observed in marine sediments³⁷, deep in the Earth's crust⁴⁶, and fascinating examples are known in which endosymbiotic methanogens cling not to free-living eubacteria, but hydrogenosomes themselves in the cytosol of amitochondriate protists^{37,43–45}. The possibility that this type of symbiotic association may have been involved in the context of mitochondrial/hydrogenosomal origins is sufficiently intriguing to examine further.

Host dependence upon hydrogen: what happens?

Let us briefly explore a hypothetical symbiosis between a free-living, H_2 - and CO_2 -producing eubacterium (the symbiont) and a methanogenic archaeebacterium (the host). They would have to meet in anaerobic environments where CO_2 and geological H_2 are abundant, so that the host is viable from the start (Fig. 2a). But once the pair is physically removed from the H_2 source (by whatever means), the host becomes immediately and strictly dependent upon the heterotrophic eubacterial symbiont (Fig. 2b). This is conceptually satisfying, because it provides a strong selective force that irreversibly associates symbiont and host. If the symbiont escapes, the host starves immediately. Such hosts are thus most successful if they (1) stick tightly to symbionts and (2) can reap the greatest benefit from them. This could conceivably select host cell shapes of large surface area that tend to surround symbionts (not endocytose them, because archaeobacteria have no cytoskeleton), increasing contact, so that more H_2 and CO_2 could be filtered through the host's cytosol (Fig. 1).

As long as the symbiont finds sufficient organic substrates, this symbiosis of prokaryotes is indefinitely sustainable, but a limitation becomes evident. Host benefit from increased surface area to the symbiont concomitantly decreases the latter's ability to effervesce gaseous life into its host, because surface contact to the environment for fuelling its own metabolism (and producing hydrogen to fuel the host) is impaired. If competition for organic substrates occurs, so will selection for hosts that find a means of utilizing their own environmental surface to import fermentable organic substrates (something that contemporary methanogens cannot do^{34,40}) for the symbiont. This could be done by evolutionary invention of

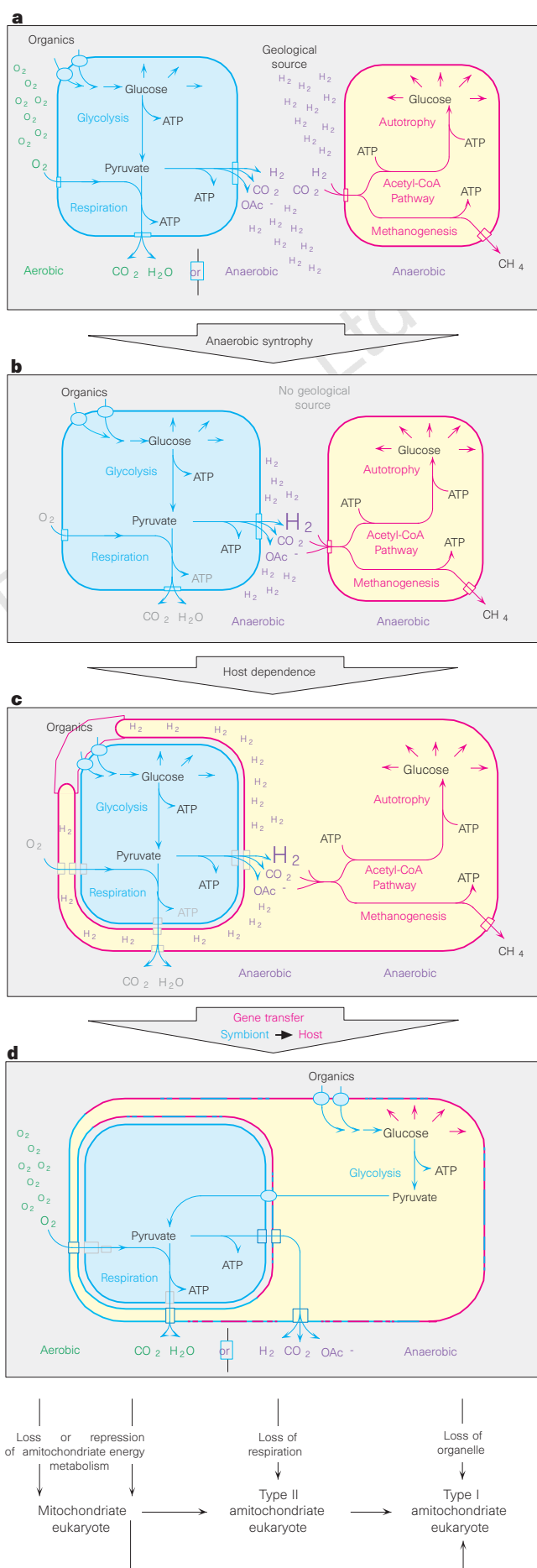


Figure 2 Hypothetical model to derive the ancestral state of eukaryotic energy metabolism put forward here, invoking strict dependence of the host upon waste products of the symbiont's anaerobic heterotrophy (see text). Host components are shaded red (cytosol yellow), symbiont components are shaded blue. The composite nature of membranes in **d** symbolizes the process of replacement of archaeobacterial lipids (glycerol ethers of isoprenes) with eubacterial lipids (glycerol esters of fatty acids) through loss of the host's lipid biosynthetic pathway. For an alternative explanation of the origin of eubacterial lipids in eukaryotes, see ref. 6. Anaerobic substrates and end products are indicated in purple, aerobic substrates and end products are indicated in green. Substrates and end products, the non-availability of which for a given step underly ecological factors, are indicated in grey.

something that did not exist before the initial meeting (importers of reduced carbon in methanogens), or without invention, by merely genetically rearranging pre-existing components. If eubacterial genes for the symbiont's carbon importers are transferred by whatever mechanism to the archaeobacterial chromosomes of the host's cytosol, are expressed there, and if the products are functional in the archaeobacterial membrane, then the host would in principle be able to feed its symbiont with organics and thus feed itself with H_2 and CO_2 (and acetate, depending upon the capability of the host). This is neither outrageously improbable, nor does it involve an evolutionary invention. It merely requires the genetic systems of eubacterium and archaeobacterium to be sufficiently compatible as to allow expression of the transferred gene(s). Such genetic compatibility may be less today than it was two or three billion years ago, at which time symbiont and host may have shared a common ancestor only one or two billion years before. Furthermore, the type of endosymbiotic gene transfer invoked here, that is, without return of the gene product to the cell compartment that donated the gene, is well documented among contemporary eukaryotes^{37,47}.

But importers alone do not allow the host to feed its symbiont. This is because, in contrast to the heterotrophic metabolism of the symbiont that generates ATP from carbohydrates, the autotrophic metabolism of the host is specialized towards synthesizing carbohydrates from CO_2 at the expense of ATP gained by other means^{34,40}. As a consequence, imported carbon flows in the wrong direction: host and symbiont alike will starve unless carbon flux to the symbiont is established, providing selection for the latter to occur. To achieve this, either (1) the host's carbohydrate metabolism must acquire, step-by-step, the regulatory properties necessary to make it run backwards (a series of evolutionary inventions), or (2) the symbiont's carbohydrate metabolism must simply be transferred to the cytosol, again through straight endosymbiotic gene transfer (single-step relocation of pre-existing components). This still does not completely solve the problem, because two pathways of carbon metabolism are now running in opposite directions (catabolic and anabolic) in the same cytosol. The result is futile cycling ($glucose + ADP \rightarrow C \text{ compounds} + ATP \rightarrow glucose + ADP$), and selection dictates that one of these pathways must be eliminated. But only if the host's pathway is eliminated can the symbiosis survive (Fig. 2d).

This leads to a curious situation. The selective pressure that associated the partners from the start and that drove the integration of eubacterial genes into archaeobacterial chromosomes was the host's strict dependence upon hydrogen produced by the symbiont. But by transferring the symbiont's importers and glycolysis to the cytosol in order to satisfy that dependence, the host suddenly can meet both its carbon and energy needs from organic substrates. The functions of both methanogenesis and autotrophy have been replaced, and there is no obvious selective pressure to retain either. The host has irreversibly become heterotrophic, and hydrogen is once again a waste product, but now of a compartmentalized metabolism.

Quite surprisingly, the result of this effortless metabolic endeavour is a hydrogenosome with a genome in an archaeobacterial host with cytosolic chromosomes, a cell that is organized in a manner strikingly similar to the amitochondriate eukaryote *Trichomonas vaginalis*¹. That this hypothetical primitive eukaryote does not possess a nuclear membrane is not disturbing; the hydrogen hypothesis simply derives a different stage of the eukaryotic cell cycle (open mitosis) than previous hypotheses do. Not a single evolutionary invention was necessary to deduce this organelle-bearing cell.

One principle, two symbioses: O_2 and plastids

Once the host's metabolic dependence upon hydrogen vanished, so would its confinement to anaerobic habitats. As outlined in previous sections, the symbiont would also have been able to respire by

virtue of its pre-existing metabolic diversity, hence O_2 could have then become advantageous to the eukaryote through respiratory ATP synthesis (and through the invention of an ATP transporter to allocate it to the cytosol). But importantly, utilization of O_2 would have been a true advantage, not a dependence as in the case of H_2 suggested here to have established the heterotrophic organelle. Such O_2 -utilization could have occurred with global increases of atmospheric O_2 levels roughly 2 billion years ago⁴⁸, or conceivably could have entailed syntrophy of the eukaryote with O_2 -producing cyanobacteria, or both, but probably in independent lineages of a young but diversified eukaryotic kingdom.

The foregoing entails the assumption that the symbiont's respiratory machinery (Krebs cycle and oxidative phosphorylation) was not lost during the H_2 -dependent phases of symbiosis, and hence the assumption that it was maintained by selection. To account for this, we suggest that the respiratory pathway of the symbiont might have enabled the anaerobic cell to free its environment of oxygen, as contemporary amitochondriate protists do (albeit by other means¹⁹). It is less evident why genes for proteins specific to energy metabolism in amitochondriate protists, such as cytosolic and hydrogenosomal PFO, have been preserved throughout the evolution of mitochondriate eukaryotes, so as to be readily recruitable in multiple lineages during the reversion to anaerobic energy metabolism, as has been clearly demonstrated among the anaerobic ciliates⁴⁴. To account for this, we offer that such enzymes perform(ed) additional essential functions for eukaryotic cells, the biochemistry of which is unknown. In support of this view are the findings that (1) dual functions for PFO in eubacteria exist, where it is an integral component of the nitrogen fixation machinery³⁰ (*nif*) by virtue of its powerful electron-donating potential, and (2) that a homologue of PFO is encoded within the yeast genome. In eukaryotes, no dual function for PFO is yet known, although it does substitute for PDH in some mitochondria, for example in *Euglena*³⁰.

The notion that syntrophy may have associated a cyanobacterial symbiont with its heterotrophic, eukaryotic host is attractive, because it would entail simple reiteration of the same principle suggested for the origin of the heterotrophic symbiont, and only the beneficial waste product of the symbiont's metabolism (O_2) is different. Thus, the contemporary benefits that both mitochondria (respiration) and plastids (photoautotrophy) confer upon their hosts may be very different and much more complex than the benefit initially provided by either (waste H_2 and O_2). A similar grade, in which biological complexity is born of chemical simplicity, has been suggested for the evolution of metabolism itself¹⁷. However, the hypothesis that one organelle may have arisen through syntrophic association does not bear on views concerning the origin of the other: the reasoning is similar but the specific premises are independent.

Conclusion

The hydrogen hypothesis can readily account for the origins of eukaryotic energy metabolism by invoking differential loss from an explicitly derived ancestral state (Fig. 2d). In doing so, it furthermore accounts for the origin of a basic eukaryotic cell in a manner that differs substantially from previous views on the topic. First, this hypothesis posits that the origins of the heterotrophic organelle (the symbiont) and the origins of the eukaryotic lineage are identical. Second, it demands only three properties of the host: (1) that it was anaerobic, (2) that it possessed strictly hydrogen-dependent metabolism and (3) that it was strictly autotrophic. It does not require the host to have possessed either nucleus, cytoskeleton, endocytosis or mitosis, therefore no organizational cline in the host lineage before the acquisition of the symbiont must be postulated. Third, it specifically posits a lethal selective force that irreversibly binds one symbiotic partner to the other. Hydrogen is the key. It is the bond that forges eukaryotes out of prokaryotes.

The archaeobacterial nature of the eukaryotic genetic apparatus

and the eubacterial nature of eukaryotic energy metabolism are premises that can be explained, not predictions that are fulfilled under this hypothesis. For both apparatuses, some exceptions to the rule can be expected—archaeobacterial transketolase in some eukaryotes³⁶ may be an example—and there is no obvious reason to expect either a eubacterial or an archaeobacterial origin for intermediate eukaryotic metabolism. Our hypothesis does not explain fundamental differences in prokaryotic and eukaryotic genome organization^{5,49}, and it does not explain the origin of eukaryotic structures that have been the focus of previous views. We also stress that the hypothetical process outlined in Fig. 2 in no way precludes the possibility that the host may have possessed a cytoskeleton before its association with the symbiont. However, we posit that the host was autotrophic: the selective advantages conferred by a cytoskeleton—arguably a prerequisite for phagocytotic feeding⁶—are less evident for an autotroph than they are for the compartmentalized heterotroph inferred here. That cell has time, energy and ample genetic starting material (two highly divergent and partially merged prokaryotic genomes) to evolve cytological and genetic traits that are specific to the eukaryotic lineage.

A methanogenic ancestry of the host is only one of several possible H₂-dependent scenarios. One in which an autotrophic host used H₂ as an electron donor, but electron acceptors other than CO₂ (sulphurous compounds, for example³⁴) could be elaborated by the same logic, whereby the host so deduced would also have been dependent upon such compounds, rather than solely upon its heterotrophic symbiont. Yet methanogenesis is attractive as the host's metabolism for several reasons. (1) It can be traced sufficiently deep into archaeobacterial phylogeny⁵⁰ as to be a candidate for a pathway ancestral to the kingdom. (2) No methanogen is known that is heterotrophic; those that utilize acetate and/or reduced C₁ compounds do so for methanogenesis and autotrophy^{34,40}. (3) Methanogens are strictly anaerobic, and (4) can utilize all three products^{34,37,40} of the symbiont's anaerobic metabolism. (5) Widespread syntrophic association between methanogens and hydrogenosomes is observable^{37,43–45}. By the criteria of simplicity under competing alternatives and of explaining unknowns in terms of known quantities, methanogenesis fares well under Occam's razor.

This hypothesis generates numerous testable predictions. We firmly predict that evidence for a strictly H₂-dependent ancestry, and most probably a methanogenic ancestry of the host should ultimately be revealed by comparative genomics. In photosynthetic eukaryotes, we predict that fewer genes of archaeobacterial ancestry should be observable, because an additional eubacterial genome is incorporated into the cell, allowing endosymbiotic gene replacement further opportunity to eliminate functionally redundant, pre-existing archaeobacterial homologues³⁶. Finally, we predict that anaerobic heterotrophic habitats devoid of geological hydrogen may harbour eukaryotes more primitive than known forms, the metabolism of which should be accountable for under the premises stated here. □

- Müller, M. The hydrogenosome. *J. Gen. Microbiol.* **139**, 2879–2889 (1993).
- Cavalier-Smith, T. Eukaryotes with no mitochondria. *Nature* **326**, 332–333 (1987).
- Sogin, M. L., Silberman, J. D., Hinkle, G. & Morrison, H. G. Problems with molecular diversity in the Eukarya. *Symp. Soc. Gen. Microbiol.* **54**, 167–184 (1996).
- Whately, J. M., John, P. & Whately, F. R. From extracellular to intracellular: the establishment of mitochondria and chloroplasts. *Proc. R. Soc. Lond. B* **204**, 165–187 (1979).
- Gray, M. W. & Doolittle, W. F. Has the endosymbiont hypothesis been proven? *Microbiol. Rev.* **46**, 1–42 (1982).
- Cavalier-Smith, T. The origin of eukaryote and archaeobacterial cells. *Ann. NY Acad. Sci.* **503**, 7–54 (1987).
- Cavalier-Smith, T. & Chao, E. E. Molecular phylogeny of the free-living archaeozoan *Trepomonas agilis* and the nature of the first eukaryote. *J. Mol. Evol.* **43**, 551–562 (1996).
- Zillig, W. *et al.* Did eukaryotes originate by a fusion event? *Endocytobiosis Cell Res.* **6**, 1–25 (1989).
- Gupta, R. S. & Golding, G. B. The origin of the eukaryotic cell. *Trends Biochem. Sci.* **21**, 166–171 (1996).
- Lake, J. A. & Rivera, M. C. Was the nucleus the first endosymbiont? *Proc. Natl Acad. Sci. USA* **91**, 2880–2881 (1994).
- Clark, C. G. & Roger, A. J. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. *Proc. Natl Acad. Sci. USA* **92**, 6518–6521 (1995).

- Henze, K. *et al.* A nuclear gene of eubacterial origin in *Euglena* reflects cryptic endosymbioses during protist evolution. *Proc. Natl Acad. Sci. USA* **92**, 9122–9126 (1995).
- Keeling, P. W. & Doolittle, W. F. Evidence that eukaryotic triosephosphate isomerase is of alpha-proteobacterial origin. *Proc. Natl Acad. Sci. USA* **94**, 1270–1275 (1997).
- Doolittle, W. F. Some aspects of the biology of cells and their possible evolutionary significance. *Symp. Soc. Gen. Microbiol.* **54**, 1–21 (1996).
- Rosenthal, B. *et al.* Evidence for the bacterial origin of genes encoding fermentation enzymes of the amitochondriate protozoan parasite *Entamoeba histolytica*. *J. Bacteriol.* **179**, 3736–3745 (1997).
- Searcy, D. G. in *The Origin and Evolution of the Cell* (eds Hartman, H. & Matsuno, K.) 47–78 (World Scientific, Singapore, 1992).
- de Duve, C. *Blueprint for a Cell: the Nature and Origin of Life* (Patterson, Burlington, NC, 1991).
- Coombs, G. H. & Müller, M. in *Biochemistry and Molecular Biology of Parasites* (eds Marr, J. J. & Müller, M.) 33–47 (Academic, London, 1995).
- Müller, M. in *Evolutionary Relationships Among Protozoa* (eds Coombs, G. H., Vickermann, K., Sleigh, M. A. & Warren, A.) 109–132 (Chapman Hall, London, 1998).
- Müller, M. Energy metabolism of protozoa without mitochondria. *Annu. Rev. Microbiol.* **42**, 465–488 (1988).
- Müller, M. in *Christian Gottfried Ehrenberg-Festschrift anlässlich der 14. Wissenschaftlichen Jahrestagung der Deutschen Gesellschaft für Protozoologie*, 9–11. März 1995 in Delitzsch (Sachsen) (eds Schlegel, M. & Hausmann, K.) 63–76 (Leipziger Universitätsverlag, Leipzig, 1996).
- Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S. & Miyata, T. Evolutionary relationship of archaeobacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl Acad. Sci. USA* **86**, 9355–9359 (1989).
- Woese, C., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eukarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
- Langer, D., Hain, J., Thuriaux, P. & Zillig, W. Transcription in Archaea: similarity to that in Eukarya. *Proc. Natl Acad. Sci. USA* **92**, 5768–5772 (1995).
- Yamamoto, A., Hashimoto, T., Asaga, E., Hasegawa, M. & Goto, N. Phylogenetic position of the mitochondrion-lacking protozoan *Trichomonas tenax*, based on amino acid sequences of elongation factors 1- α and 2. *J. Mol. Evol.* **44**, 98–105 (1997).
- Horner, D. S., Hirt, R. P., Kilvington, S., Lloyd, D. & Embley, T. M. Molecular data suggest an early acquisition of the mitochondrion endosymbiont. *Proc. R. Soc. Lond. B* **263**, 1053–1059 (1996).
- Bui, E. T. N., Bradley, P. J. & Johnson, P. J. A common evolutionary origin for mitochondria and hydrogenosomes. *Proc. Natl Acad. Sci. USA* **93**, 9651–9656 (1996).
- Germot, A., Philippe, H. & Le Guyader, H. Presence of a mitochondrial-type 70-kDa heat shock protein in *Trichomonas vaginalis* suggests a very early mitochondrial endosymbiosis in eukaryotes. *Proc. Natl Acad. Sci. USA* **93**, 14614–14617 (1996).
- Roger, A. J., Clark, C. G. & Doolittle, W. F. A possible mitochondrial gene in the early-branching amitochondriate protist *Trichomonas vaginalis*. *Proc. Natl Acad. Sci. USA* **93**, 14618–14622 (1996).
- Hrdý, I. & Müller, M. Primary structure and eubacterial relationships of the pyruvate:ferredoxin oxidoreductase of the amitochondriate eukaryote, *Trichomonas vaginalis*. *J. Mol. Evol.* **41**, 388–396 (1995).
- Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
- Kaneko, T. *et al.* Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109–136 (1996).
- Sánchez, L. B. & Müller, M. Purification and characterization of the acetate forming enzyme, acetyl-CoA synthetase (ADP-forming) from the amitochondriate protist, *Giardia lamblia*. *FEBS Lett.* **378**, 240–244 (1996).
- Schönheit, P. & Schäfer, T. Metabolism of hyperthermophiles. *World J. Microbiol. Biotechnol.* **11**, 26–57 (1995).
- Markoš, A., Miretsky, A. & Müller, M. A glyceraldehyde-3-phosphate dehydrogenase with eubacterial features in the amitochondriate eukaryote *Trichomonas vaginalis*. *J. Mol. Evol.* **37**, 631–643 (1993).
- Martin, W. & Schnarrenberger, C. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr. Genet.* **32**, 1–18 (1997).
- Fenchel, T. & Finlay, B. J. *Ecology and Evolution in Anoxic Worlds* (Oxford Univ. Press, Oxford, 1995).
- Gibson, J. L. & Tabita, F. R. The molecular regulation of the reductive pentose phosphate pathway in proteobacteria and cyanobacteria. *Arch. Microbiol.* **166**, 141–150 (1996).
- Murrel, J. C. Genetics and molecular biology of methanotrophs. *FEMS Microbiol. Lett.* **88**, 233–248 (1992).
- Thauer, R. K., Hedderich, R. & Fischer, R. in *Methanogenesis: Ecology, Physiology, Biochemistry and Genetics* (ed. Ferry, J. G.) 209–252 (Chapman & Hall, New York, 1993).
- Conrad, R. Soil microorganisms as controllers of atmospheric trace gases (H₂, CO, CH₄, OCS, N₂, and NO). *Microbiol. Rev.* **60**, 609–640 (1996).
- Bryant, M. P., Wolin, E. A., Wolin, M. J. & Wolfe, R. S. *Methanobacillus omelianskii*, a symbiotic association of two species of bacteria. *Arch. Microbiol.* **59**, 20–31 (1967).
- Broers, C. A. M., Stumm, C. K., Vogels, G. D. & Brugerolle, G. *Psalteriomonas lanterna* gen. nov., sp. nov., a free living amoboflagellate isolated from freshwater anaerobic sediments. *Eur. J. Protistol.* **25**, 369–380 (1990).
- Embley, T. M. *et al.* Multiple origins of anaerobic ciliates with hydrogenosomes within the radiation of aerobic ciliates. *Proc. R. Soc. Lond. B* **262**, 87–93 (1995).
- Brinkmann, H. & Martin, W. Higher plant chloroplast and cytosolic 3-phosphoglycerate kinases: a case of endosymbiotic gene replacement. *Plant. Mol. Biol.* **30**, 65–75 (1996).
- Kasting, J. F. Earth's early atmosphere. *Science* **259**, 920–926 (1993).
- Poole, A. M., Jeffares, D. C. & Penny, D. The path from the RNA world. *J. Mol. Evol.* **46**, 1–17 (1998).
- Rospert, S. *et al.* Methyl-coenzyme M reductase and other enzymes involved in methanogenesis from CO₂ and H₂ in the extreme thermophile *Methanopyrus kandleri*. *Arch. Microbiol.* **156**, 49–55 (1991).

Acknowledgements. We thank H. Brinkmann, M. Embley, K. Henze, R. Herrmann, R. Hensel, D. Oesterheld and L. Sánchez for critical comments on the manuscript and gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft (W.M.) and the National Institutes of Health (M.M.).

Correspondence should be addressed to W.M. (e-mail: w.martin@tu-bs.de) or to M.M. (e-mail: mmuller@rockvax.rockefeller.edu).

Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA

Lin Chen*, J. N. Mark Glover*†, Patrick G. Hogan‡, Anjana Rao‡ & Stephen C. Harrison*†

* Department of Molecular and Cellular Biology and † Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA

‡ Center for Blood Research, Harvard Medical School, Boston, Massachusetts 02115, USA

The nuclear factor of activated T cells (NFAT) and the AP-1 heterodimer, Fos–Jun, cooperatively bind a composite DNA site and synergistically activate the expression of many immune-response genes. A 2.7-Å-resolution crystal structure of the DNA-binding domains of NFAT, Fos and Jun, in a quaternary complex with a DNA fragment containing the distal antigen-receptor response element from the interleukin-2 gene promoter, shows an extended interface between NFAT and AP-1, facilitated by the bending of Fos and DNA. The tight association of the three proteins on DNA creates a continuous groove for the recognition of 15 base pairs.

Engagement of the T-cell antigen receptor (TCR) by presented antigen induces a cascade of transcriptional responses^{1,2}. The synthesis of interleukin (IL)-2, an important early event, depends on a regulatory region that extends about 300 base pairs upstream of the transcriptional start site, and several signalling pathways, from the TCR and from co-stimulatory receptors, converge on this promoter^{3,4}. The transcription factor known as nuclear factor of activated T cells (NFAT) binds to several sites within the regulatory region of the IL-2 gene and other inducible genes^{5,6}. Its properties have been analysed in particular detail, because its activation is inhibited by the main immunosuppressive drugs in current clinical use, cyclosporin A and FK506 (ref. 7). There are four known mammalian NFAT genes^{8–11}, which encode proteins of about 700–1,100 amino-acid residues. There are two contiguous, well-conserved regions—the so-called NFAT homology region (NHR, 300 residues) and the DNA-binding region (about 270 residues)—sandwiched between more variable ‘transactivation domains’ at either end of the polypeptide chain⁶. The DNA-binding region bears a distant sequence similarity to the Rel homology region (RHR) of Rel-family transcription factors such as NF-κB^{12,13}.

In resting T cells, NFAT proteins are largely cytoplasmic. Stimulation of the TCR leads to increased intracellular Ca²⁺, dephosphorylation of sites in the NHR by calcineurin (a calcium-activated phosphatase), and ultimately nuclear import and increased DNA binding of NFAT^{14–18}. These steps are blocked by the calcineurin inhibitors CsA and FK506^{7,14}. Full response at many NFAT sites requires concomitant activation of members of the AP-1 transcription-factor family^{6,20} (the AP-1 partner is referred to as the ‘nuclear component’ of NFAT in much of the original literature on this factor)^{1,2}. There is a discernible AP-1 (Fos–Jun) site immediately downstream of most NFAT sites in the promoter region of IL-2 and other cytokine genes, but it usually differs at more than one position from the consensus AP-1 sequence⁶. Indeed, relatively weak binding at their respective sites by NFAT and AP-1 individually, together with significant cooperative interaction between the two on composite sites, reflects the functional requirement for both transcription factors^{21,22}.

The DNA-binding domains of NFAT (the RHR) and AP-1 (the bZIP elements of the Fos and Jun partners) are necessary and sufficient for cooperative association on DNA²³. Thus interactions between the adjacently bound proteins must be sufficiently favourable that cooperative binding (and the functional counterpart, synergistic activation) can occur at concentrations below the K_d of NFAT or AP-1 alone. To see directly the contact that two such

unrelated structures might make, and to work out the mechanism for this part of the convergent, multipathway T-cell response, we have crystallized a quaternary complex containing the bZIP elements of Fos and Jun, the RHR of human NFAT1 (ref. 12), and a DNA fragment that has the sequence of the distal antigen-receptor response element (ARRE2) from the murine IL-2 promoter²³. The structure reveals an extensive interface between NFAT and AP-1, facilitated by bending of the Fos helix and the DNA. The DNA-binding surfaces of NFAT and AP-1 together form a continuous groove for the recognition of 15 base pairs.

Overview of the structure

The NFAT DNA-binding region (residues 399–678) and the AP-1 bZIP elements form a tight complex with each other and with the DNA (Figs 1 and 2). The NFAT fragment is very similar in the folded structures of its domains to the RHR of NF-κB p50 (refs 24, 25; see also the published nuclear magnetic resonance (NMR) structure of the amino-terminal domain²⁶), but the conformation of the linker is quite different, so that the cleft between the two domains closes up around the Fos–Jun heterodimer. Two structural adaptations ensure that AP-1 has an extended interaction with NFAT (Fig. 3a, b): the coiled coil inclines by about 15° toward the 5′ end of the site, and the DNA itself bends by about 20° at the centre of the AP-1 site. The Fos–Jun heterodimer is oriented specifically so that Jun binds to the half of the AP-1 site close to NFAT, as demonstrated in studies of the complex in solution²⁷.

NFAT

The two domains of the NFAT DNA-binding region identify it clearly as a Rel homologue, despite the lack of extensive sequence conservation (Fig. 1a). Both domains have immunoglobulin-like folds.

The N-terminal domain contains, in addition to the strands of a ‘standard’ s-type immunoglobulin module²⁸, two additional strands between C and C′. We designate these strands X and Y, to preserve the conventional immunoglobulin-strand nomenclature for the rest of the domain. The ABE sheet and the GFCC′Y sheet form the two faces of the β-sandwich, and the two-strand sheet XE′ and the AB loop seal the bottom. The E′F loop contains a short α-helix. In p50, this loop is the site of an approximately 68-residue, largely α-helical ‘insert’ into the minimal RHR. In NFAT, the loop is even shorter (21 residues) than in p65 or c-Rel, and forms an important part of the interface between NFAT and AP-1.

The carboxy-terminal domain has the h-type immunoglobulin

Figure 1 Amino-acid sequences of NFAT, NF- κ B p50 and the bZIP regions of Fos and Jun. **a**, Amino-acid sequences of the DNA-binding regions of the four NFAT family members (human) and of NF- κ B p50. The alignment with NF- κ B (p50) is based on the structures. Secondary-structure assignments from our model of NFAT1 are shown as coloured bars (α -helices) and arrows (β -strands) above the aligned sequences. Broken lines denote disordered regions. Our lettering adheres essentially that of ref. 25, except that in the N-terminal domain we have used X and Y instead of H and I, to emphasize that these strands are insertions into the basic immunoglobulin design. The p50 lettering of ref. 24 simply assigned alphabetically sequential letters to the strands, thus departing from immunoglobulin convention after strand C. We suggest that the present convention be used for this entire superfamily of transcription factors. The colour code for the N-terminal domain is yellow and for the C-terminal domain green. The coloured blocks show residues that participate in contacts to DNA (magenta), Fos (red) or Jun (blue) or in interdomain contacts within NFAT (light green). Asterisks indicate variable contacting residues among family members. **b**, **c**, Amino-acid sequences of the bZIP regions of Fos (**b**) and Jun (**c**) family members (human). Coloured blocks show contacts with DNA (magenta) or with NFAT (yellow or green, according to the domain contacted). Asterisks indicate variable contacting residues among family members.

fold²⁸. It folds back against the N-terminal domain (Fig. 2). The corresponding domain of Rel family members, which it closely resembles, would direct dimerization, but NFAT is clearly a monomer both in solution^{11,27} and in this complex. The outer surface of the abed sheet is not significantly more polar in NFAT, but it is oriented in such a way that a partner with a Rel-like dimer contact could not bind DNA without major reconfiguration of the inter-domain linker. This linker includes an α -helical segment that abuts the DNA backbone and contributes to base-pair recognition, and a linker rearrangement would therefore perturb both base-pair and backbone contacts. In the unbound state, however, it is likely that the linker is flexible. The inter-domain contacts do not appear sufficiently tight to maintain the relative orientation of the domains in the free protein. They may, however, contribute to the stability of the particular domain arrangement seen in the DNA-bound structure.

AP-1

The segments of Fos and Jun are each single α -helices, and the heterodimer forks across the DNA major groove, just as in the DNA complexes of other bZIP proteins^{29–31}. The asymmetry of the coiled coil that was seen in our earlier crystal structure of a Fos–Jun–DNA complex³¹ is evident here as well. The leucine-zipper part of Jun curls markedly around the straighter zipper element of Fos. The coiled coils in the two structures superpose well ($C\alpha$ root-mean-square deviation 0.42 Å), showing that the asymmetry is a property of the Fos–Jun contacts and not a consequence of interactions with NFAT or other molecules in the crystal lattice. The Fos–Jun–DNA crystals studied previously contained two distinct complexes in the crystallographic asymmetric unit. These differed not only in their preferred orientation for the heterodimer on the DNA site but also in the way the two α -helices were bent in the fork between the basic regions and the coiled coil. In this quaternary complex with NFAT (Fig. 3), the entire Jun segment in the present structure can be superposed, together with the leucine zipper of Fos, on the corresponding part of one of the two complexes in the earlier structure (Fig. 3c), but the fork region of Fos flexes by about 15° around a pivot near Glu 160, so the axis of the coiled coil leans markedly away from a perpendicular to the axis of the DNA (Fig. 3a, b). Examination of the crystal packing shows that lattice contacts cannot be the cause of the distortion. Differences in the AP-1 site DNA conformation between Fos–Jun complexes with and without NFAT allow the local major-groove interactions of Fos and Jun to remain similar,

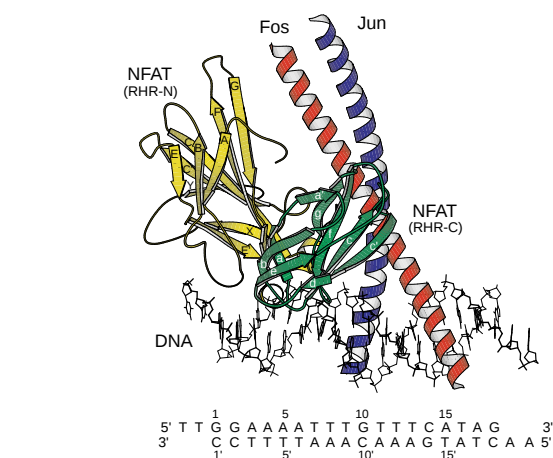
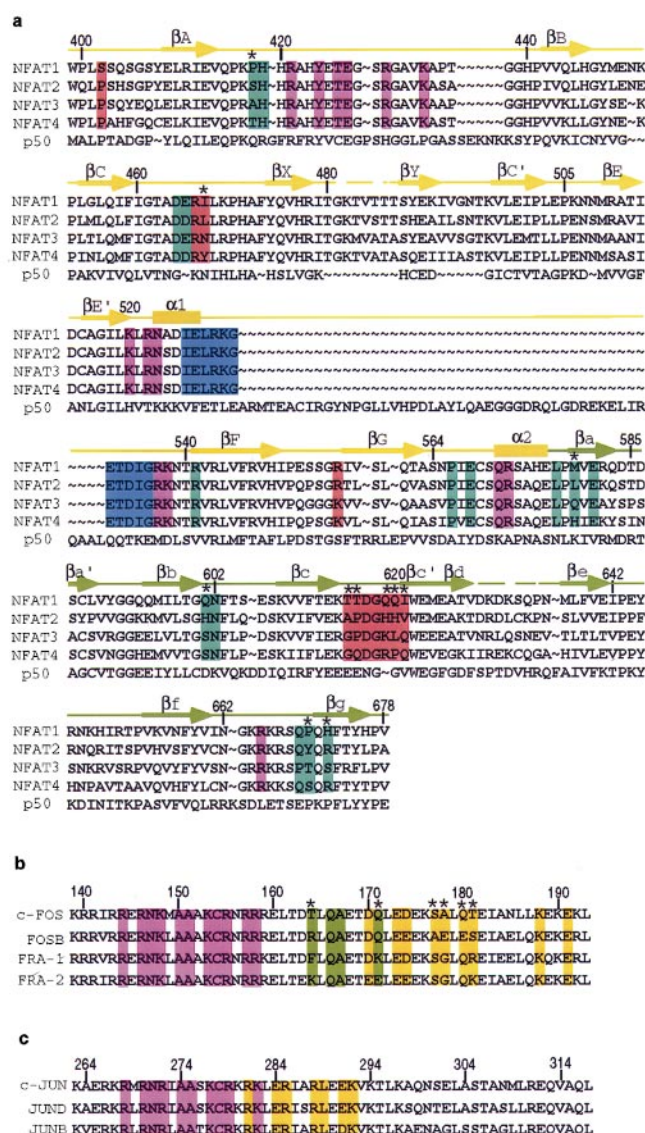


Figure 2 The NFAT-AP-1-DNA complex (drawn with MOLSCRIPT⁴⁰). The colours here are used throughout the rest of the illustrations: the N-terminal domain (RHR-N) of the NFAT RHR is yellow; the C-terminal domain (RHR-C) is green; Fos is red; Jun is blue. Strands and helices are labelled; the lettering corresponds to that used for the secondary-structure indicators in Fig. 1. The ARRE2 nucleotide sequence in the crystals is shown below the figure, together with the numbering scheme used in the text.

even after such a significant adjustment in one of the partners (see below).

DNA conformation in the complex

The entire DNA conforms to a B-type helix. The complementary two-base overhang at either end of the fragment establishes continuity of the double helix from one unit cell to the next. There is a bend of about 20° at the centre of the AP-1 site, towards the major groove and hence towards the coiled-coil side of the Fos–Jun heterodimer (Fig. 3a). Rather than narrowing the major groove, however, this bend appears to have the effect of displacing the sugar-phosphate backbone on the Jun side of the complex towards NFAT, so the minor groove in the vicinity of phosphate P7 is narrowed instead. This narrowing extends over the entire A:T-rich segment from base pairs 3 to 8. The reverse bend, required for continuity of DNA in the crystal, occurs in the base-paired overhang.

DNA recognition

Overall. The combined DNA-binding surfaces of NFAT, Fos and Jun contact the extended ARRE2 composite site continuously (Fig. 4). Bases in the core NFAT site (GGAAAA) and the AP-1 site (TGTTTCA) interact with NFAT and Fos–Jun, respectively, in the major groove, and the intervening spacer sequence (TT) interacts with an NFAT arginine residue in the minor groove (Fig. 4b, c).

NFAT recognition. The AB loop and the α -helical segment of the linker supply side chains that contact bases in the major groove. The AB loop is the homologue of the 'recognition loop' in NF- κ B p50 (refs 24–26), and its conformation and contacts closely resemble those of its p50 counterpart (Fig. 4b, c). Arg 430 and Arg 421 participate in bidentate hydrogen-bond interactions with O6 and N7 of Gua 1 and Gua 2, respectively. The guanidinium groups of the arginines stack upon each other, following the stack of the corresponding guanine bases. In NFAT, the side-chain amide group of Gln 571 extends this stack and engages in bidentate hydrogen-bonding with Ade 3 (refs 32, 33). The GGA sequence thus recognized seems to be the essential element of an NFAT site. The remaining A:T base pairs in the GGAAAA sequence, which appears to be well conserved among many NFAT sites⁶, are specified by van

der Waals contacts with their thymine methyl groups. The ring of Tyr 424 is the partner for thymines 3' and 4'; the aliphatic part of Arg 572 is the partner for thymines 5' and 6'. Both residues are also anchored by hydrogen bonds to the DNA backbone.

The E'F loop bridges across the DNA backbone from the major to the minor groove. Arg 537, at the end of this loop, dips into the minor groove and donates a hydrogen bond to the carbonyl of Thy 8. The segment of DNA backbone between Thy 3' and Ade 7' is sandwiched by the linker α -helix in the major groove and the E'F loop in the minor groove and by the AB loop at one end and the fg loop at the other end. A strong ethylation interference footprint covers this region³⁴.

AP-1 recognition. The seven base-pair AP-1 site in the ARRE2 deviates from consensus at two positions, but many of the interactions between basic-region side chains and base-pair functional groups are the same as those found in the ternary Fos–Jun–DNA complexes³¹ (Fig. 4b). The bend in the fork segment of Fos, the invariance of the zipper interactions between Fos and Jun, and the overall invariance of the Jun conformation might then have been expected to generate a displacement of the Jun basic region with respect to the DNA major groove. But underwinding and bending at the centre of the AP-1 site together restore the register of base pairs with residues in Jun (Fig. 4b), even though the orientation of the Jun basic region in the major groove is nearly perpendicular to the DNA axis. The conserved asparagine residue in the Jun basic region (Asn 271) accepts at least one hydrogen bond (from N4 of Cyt 10') present in all the other GCN4 and Fos–Jun complexes, and Ala 274 also has a conserved contact with the methyl group of Thy 9.

Comparison with biochemical and mutational data. The base-pair and sugar-phosphate contacts shown in Fig. 4b readily account for most of the observed effects of chemical modifications in the DNA and mutational substitution in NFAT^{12,26,34,35}. Methylation interference from Ade 3 to Ade 8' can probably be ascribed to the minor-groove compression seen in the structure and to the contact between Arg 537 and Thy 8. Because the interference is seen also in the binary NFAT–DNA complex, the compression may be present even with NFAT alone. As explained above, minor groove narrowing

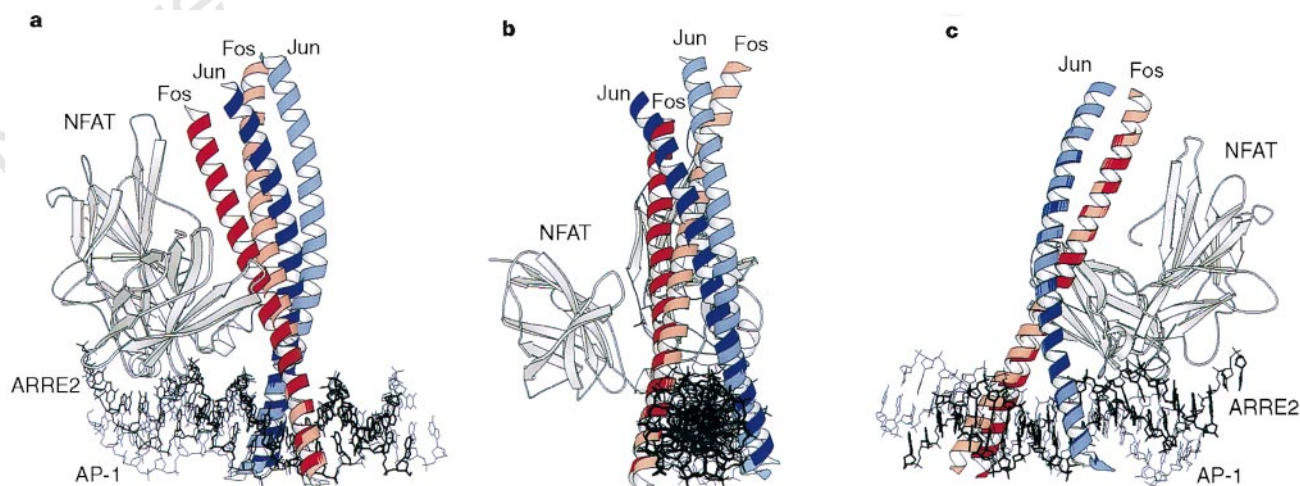


Figure 3 Superpositions of the NFAT–Fos–Jun–ARRE2 quaternary complex on the Fos–Jun–AP-1 site ternary complex (complex I of ref. 31). These comparisons (drawn with MOLSCRIPT) illustrate that bending of Fos and of DNA together allow the formation of the observed NFAT–Fos–Jun interface. **a**, The ternary (Fos–Jun–AP-1 site) and quaternary (NFAT–Fos–Jun–ARRE2) complexes have been superposed so that the basic region of Fos and the segment of DNA that it contacts are optimally aligned. Lighter colours and thin lines indicate the ternary complex³¹; darker colours and thick lines show the quaternary complex; NFAT is shown grey.

The DNA bend can be seen by comparing the two superposed structures. **b**, End-on view of the same superposition. **c**, The two structures have been superposed so that the leucine zippers are optimally aligned. This superposition, when compared with the superposition used in **a** and **b**, shows that when shifting from the conformation in the ternary complex to that in the quaternary complex, the rest of the Fos–Jun–DNA complex can be imagined to pivot about the fork segment of Fos.

accompanies DNA bending, which is in turn essential for the interaction between AP-1 and NFAT (see below). The narrow minor groove is probably stabilized in part by insertion of Arg 537 opposite Thy 8 and Arg 665 between Ade 7' and Gua 10. Various NFAT-AP-1 composite sites exhibit a strong bias between positions 4 and 8 for A:T or T:A base pairs, and minor-groove narrowing occurs more readily in A:T-rich sequences. This apparent 'indirect' specification of A:T or T:A between base pairs 4 and 8 complements the 'direct' preference for T at 4', 5', and 6', owing to van der Waals contacts from Tyr 424 and Arg 572.

Different NFAT family members have been reported to have distinct DNA-binding specificity and affinity⁶. The crystal structure

shows, however, that residues involved in DNA recognition are highly conserved among family members (Fig. 1a).

The interface between NFAT and AP-1

The N-terminal domain of the NFAT RHR has a large contact surface with Fos and Jun, extending over much of the length of the fork and the coiled coil and merging with the protein–DNA interface (Fig. 5). The NFAT C-terminal domain also has a small contact with Fos. The interactions are primarily polar, creating a network of hydrogen bonds and salt links that connect Fos, Jun, NFAT and the DNA backbone (Fig. 5b). The contacts of the E'F loop with Jun and of the CX loop and N-terminal segment with Fos are each centred

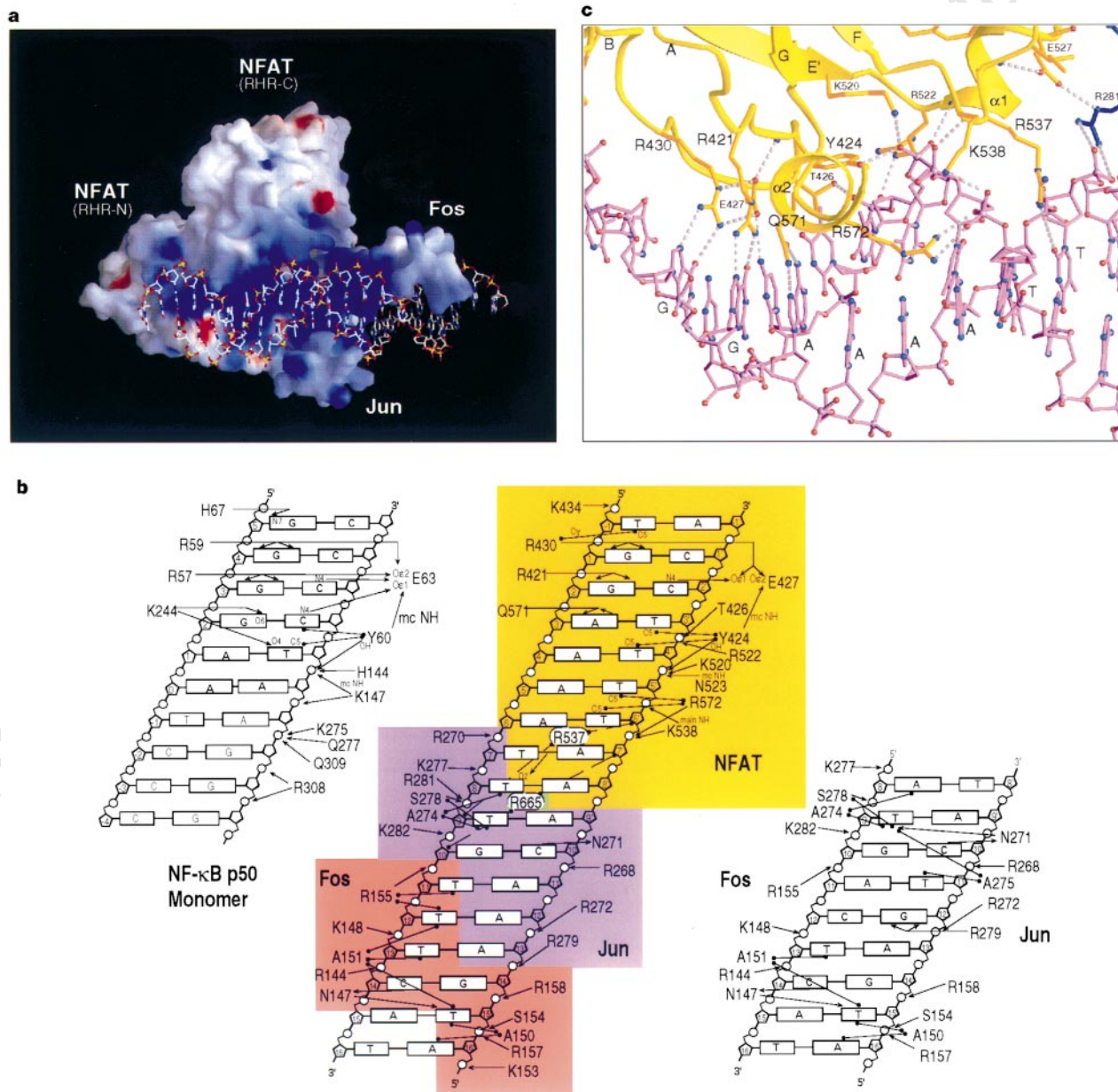


Figure 4 DNA recognition in the NFAT-AP-1-ARRE2 complex. **a**, View of the composite binding groove generated by the DNA-facing surfaces of NFAT, Fos and Jun. The colour scheme indicates electrostatic potential (blue, positive; red, negative) as calculated by the program GRASP⁴⁹. **b**, Schematic diagram of detailed protein-DNA contacts, compared with similar contacts in the NF- κ B p50-DNA complex (upper left)²⁴ and in the ternary Fos-Jun-AP-1 site complex (lower right)³¹. Lines with arrowheads, hydrogen bonds; lines with filled circles at

either end, van der Waals contacts; solid lines, major-groove and sugar-phosphate interactions; dashed lines, minor-groove interactions. The coloured blocks indicate which protein is involved. **c**, Detailed view (using Ribbons⁵⁰) of NFAT interactions in the major groove, showing the roles of Arg 421, Glu 427, Arg 430, Gln 571 and Arg 572 in specifying base pairs. Arg 430 is supported by salt links to Glu 427, which also accepts hydrogen bonds from the main-chain NH of Tyr 424 and from N4 of Cyt 2'. The view is essentially the same as that in Fig. 2.

on a small non-polar patch, surrounded by an extended cluster of hydrogen bonds. The hydrogen-bond networks communicate with each other, but there is not a fully continuous van der Waals complementarity of the AP-1 and NFAT surfaces between one patch and the next. That is, the extended interaction region contains small hydrated pockets between these major foci of contact (Fig. 5a).

The E'F loop contacts the fork segment of Jun close to the DNA surface. The loop extends away from a DNA phosphate (P4') as a short α -helix ($\alpha 1$), and it returns toward the adjacent phosphate (P5') as a strand (Fig. 5b). This loop appears to be a segment of structural variability in Rel-family proteins, and our structure suggests that this variability corresponds to different partners for

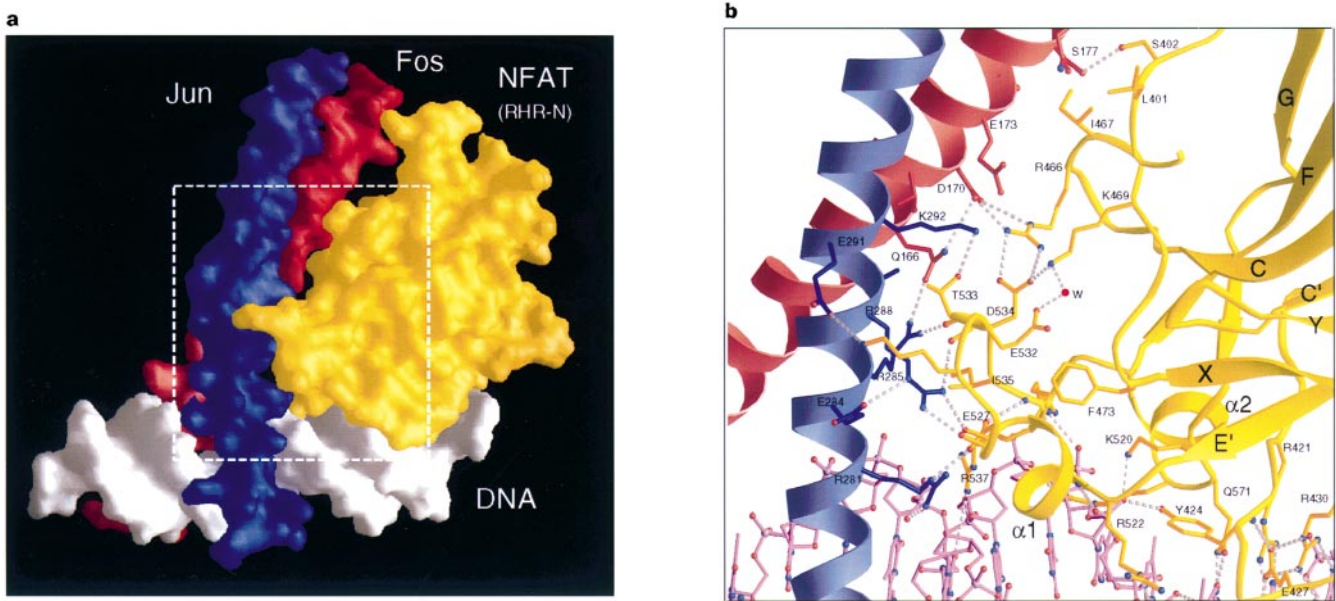


Figure 5 The interface between NFAT and AP-1. **a**, Surface representation (made with GRASP) showing the partly discontinuous character of the contacts. The box indicates the region shown in detail in **b**. The orientation is essentially the same as

in Fig. 3c. **b**, Detailed view (using Ribbons⁵⁰) of the contacts between the E'F and CX loops of NFAT (right) and Jun and Fos (left), respectively. Note the very extensive network of hydrogen bonds and salt bridges.

Table 1 Data collection, phase determination, and refinement statistics

Crystallographic data								
Data set	Cell	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	β (°)	Resolution (Å)	Coverage (redundancies)	<i>R</i> _{sym}
Native I	P2 ₁	64.76	85.55	83.37	112.41	30–2.85	90.0% (1.9)	2.5%
Native II	P2 ₁	65.19	87.23	84.10	112.66	30–2.85	96.7% (2.5)	4.6%
Native III	P2 ₁	64.55	85.46	83.37	112.00	20–2.70	98.3% (3.1)	8.0%
Phase determination (Native I as native)								
	Resolution (Å)	Cover (redun)	<i>R</i> _{sym} *†	<i>R</i> _{deriv} (sites)†	<i>R</i> _{cullis} (ano)‡	Phasing power§	f.o.m. (res)	
MIR								
Iodine 3	20–2.85	79% (2.7)	5.1%	9.2% (4)	0.79	1.27		
Iodine 5	20–2.95	89% (2.9)	5.1%	15.0% (4)	0.86	1.17		
Hg	20–3.00	92% (2.4)	8.6%	24.0% (4)	0.79	1.29		
MAD								0.32 (3.5 Å)
$\lambda 1$ (0.9650 Å)	20–3.00	90% (4.5)	9.9%	(6)	(0.95)			
$\lambda 2$ (0.9795 Å)	20–3.00	90% (4.5)	10.5%	(6)	0.93 (0.92)	0.66		
$\lambda 3$ (0.9809 Å)	20–3.00	90% (4.5)	10.2%	(6)	0.92 (0.95)	0.60		
Refinement (Native I and II, 3,887 non-hydrogen atoms; native III, 3,975 non-hydrogen atoms)								
	Resolution (Å)	Reflections (free set)	<i>R</i> _{cryst} / <i>R</i> _{free} (2 σ)¶	Bonds (Å)	R.m.s.d.#	Average <i>B</i> -factor (Å ²)	Protein	DNA
Native I	10–2.85	16531 (1215)	23.6%/29.5%	0.009	1.51	57.7		46.2
Native II	10–2.85	17893 (1950)	24.0%/30.1%	0.011	1.62	60.8		48.4
Native III	10–2.7	20408 (1671)	24.6%/30.3%	0.010	1.62	54.3		39.4

Native I, II and III refer to three distinct data sets from slightly non-isomorphous crystals. Native III is the complex crystal in which NFAT is selenomethionine-substituted. Iodine 3 is a derivative in which T7, T8, T16 and T17' have been replaced by 5-iodo-dU; in iodine 5, T(–1), T7, T16 and T17' have been replaced by 5-iodo-dU. The refined structure contains 266 residues of NFAT, 53 of Fos, 52 of Jun, 20 DNA base pairs and 88 water molecules.

* $R_{sym} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity, $\langle I \rangle$ is the statistically weighted average intensity of multiple observations of symmetry-related reflections.

† $R_{deriv} = \sum ||F_{PH}| - |F_P|| / \sum |F_P|$, where $|F_P|$ is the protein structure factor amplitude and $|F_{PH}|$ is the heavy-atom derivative structure factor amplitude.

‡ $R_{cullis} = \sum ||F_{PH}| - |F_P + F_H|| / \sum |F_{PH} - F_P|$; $R_{cullis} (ano) = \sum ||F_{PH+} - F_{PH-}|_{obs} - |F_{PH+} - F_{PH-}|_{calc}| / \sum |F_{PH+} - F_{PH-}|_{obs}$, where F_H is the heavy-atom structure factor amplitude.

§ Phasing power = $\langle |F_H| \rangle \langle ||F_{PH}| - |F_P + F_H|| \rangle / \langle |F_{PH}| - |F_P + F_H| \rangle$ is also known as the residual lack of closure error.

|| f.o.m. is figure of merit = $(\sum P(\alpha) e^{i\alpha}) / \sum P(\alpha)$, where α is the phase and $P(\alpha)$ is the phase probability distribution.

¶ $R = \sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o and F_c are observed and calculated structure factor amplitudes, respectively. R_{free} is calculated for a randomly chosen 7.5% of reflections; R_{cryst} is calculated for the remaining 92.5% of reflections used for structure refinement.

R.m.s.d. is the root mean square deviation from ideal geometry.

cooperative interactions. The loop is supported by a small hydrophobic patch, centred on Phe 473. Several basic residues and two main-chain amides from the E'F loop contact the DNA backbone. The N terminus of $\alpha 1$ is capped by phosphate 4'; the C terminus is capped by Arg 288 of Jun. Thr 533 at the tip of the loop packs closely against the Jun helix, in a pocket created by the aliphatic segments of Arg 285, Arg 288 and Leu 289, and it also participates in the hydrogen-bond network. The E'F loop thus provides a continuous link at the protein interface between contacts to DNA backbone and contacts to Jun, and the presence of both DNA and Jun appear to stabilize its projecting conformation.

The CX loop, the N-terminal segment and the FG loop of NFAT contact Fos in successive 'layers' along the zipper helix. The tip of the CX loop (Arg 466 and Ile 467) interacts in the region between Asp 170 and Ser 177. The FG loop lies opposite the end of the Fos polypeptide we have used, and there is a salt link between Arg 556 and Glu 191. The cc' loop of the NFAT C-terminal domain is adjacent to residues 164 to 171 of Fos. The actual non-covalent interactions appear to be limited, however, with only one clear hydrogen bond and a small patch of van der Waals contact.

There are four known members of the NFAT family and several members each of the Fos and Jun families. Residues involved in the NFAT–Jun contacts are invariant in both families (Fig. 1). By contrast, there is some variation of sequence on the two sides of the Fos–NFAT interface, particularly at the contact with the C-terminal domain of the NFAT RHR.

Alanine scanning mutagenesis has indicated that Arg 285 of Jun and Thr 533, Glu 527 and Ile 535 of NFAT participate in cooperative DNA binding by NFAT and AP-1 (refs 35, 36). Directed mutations in Arg 466 and Ile 467 also influence the interaction between NFAT and AP-1 (F. Macian and A. Rao, unpublished results). All these residues are in the observed interface (Figs 1 and 5b). However, contacts between NFAT and AP-1 are not strong enough to stabilize the complex when not bound to its DNA site, as shown by protein crosslinking experiments in the presence and absence of specific DNA²⁷.

Combinatorial regulation of transcription

Composite sites such as ARRE2 integrate signals from several pathways⁶. Weak independent binding of NFAT or AP-1 and strong cooperativity imply that both partners must be activated to generate a transcriptional response. A noteworthy characteristic of transcription-factor cooperativity on composite sites is the direct participation of the DNA-binding domains^{23,37–39}. For example, mutations that affect interactions between AP-1 and the glucocorticoid receptor on the proliferin promoter map to outward-facing residues in the basic region of Fos³⁷. Our structure illustrates that interactions between DNA-binding domains will constrain the design and evolution of the contacting surfaces. The interface between NFAT and AP-1 is strikingly extended, and some distortions, both of the AP-1 partner and of the DNA, appear to be required to achieve the observed match. The actual cooperativity (about a 10-fold increase in affinity of AP-1 for DNA³⁶) is therefore the difference between these presumably unfavourable distortion components and favourable contributions from the various NFAT–AP-1 contacts.

Could NFAT function with partners other than AP-1? For example, could a transcription factor with a totally different sort of DNA-binding domain supply adequate cooperativity? The NFAT–AP-1 complementarity, although extensive, is partly discontinuous. There are discrete patches of interactions, and the polar side chains at the contacting surface of NFAT give it significant adaptability. It is therefore possible that a different kind of transcription factor could provide an interaction surface for a different subset of the loops on the 3' facing side of NFAT. Contacts on the opposite, 5'-facing surface (the C'E and XY loops) may also occur in a fully occupied regulatory region such as the IL-2 promoter.

There is evidence that active promoters and enhancers are essentially covered with bound factors and that their DNA may be looped into a compact, three-dimensionally ordered arrangement^{39–41}. The quaternary complex we describe here is therefore likely to be a substructure in a still larger assembly, and some further components could associate with the combined surfaces of Fos, Jun and NFAT. □

Methods

Preparation of NFAT complex. Expression and purification of recombinant human NFAT1 (396–678, with an 18-residue N-terminal His-tag), Fos (c-Fos, 139–193), and Jun (c-Jun, 263–317) are as described previously^{31,34}. Selenomethionine (SeMet)-substituted NFAT was prepared by growing NFAT-producing *Escherichia coli* in minimal medium in which methionine was replaced by SeMet. Oligonucleotides and their iodo derivatives were synthesized on a Milligen DNA synthesizer and purified by reverse-phase high-performance liquid chromatography. The complex was prepared by mixing 1.5:1.5:1.5:1 equivalents of NFAT, Fos, Jun and DNA at a concentration of 0.2–0.4 mM in a storage buffer containing 10 mM HEPES (pH 7.5), 1 mM DTT, 100 mM NaCl, 20% glycerol and 500 mM NH₄OAc.

Crystallization and data collection. Hanging drops of 1–2 μ l NFAT complex stock solution were equilibrated with 700 μ l reservoir solution that had the same chemical composition as the storage buffer, but a lower NH₄OAc concentration (300 mM). Trials with 5 NFAT RHR variants, Fos and Jun zippers of 4 different lengths, and about 60 DNA duplexes ultimately yielded crystals that diffracted to a spacing of 2.7 Å. The NFAT1, Fos and Jun sequences are derived from the human proteins (Fig. 1); the ARRE2 sequence (Fig. 2) is murine but differs at only two positions from the human ARRE2 (ref. 6). At room temperature, crystals grew to maximum size (0.4 × 0.2 × 0.03 mm³) as thin plates within 24 h. Crystals were stable in the harvest/cryoprotectant buffer: 10 mM HEPES (pH 7.5), 100 mM NaCl, 10% PEG3K, 15% glycerol, 15% ethyleneglycol. There is one complex in the asymmetric unit (68% solvent). Isomorphous heavy-atom derivatives were obtained with iodinated DNA (Table 1) and by soaking crystals in methyl mercury nitrate. All native and derivative crystals were flash frozen in liquid nitrogen for storage and data collection under cryogenic conditions (100 K).

Data were collected with a Mar Research image plate scanner at NSLS (beamline X-25) and at SSRL (beamline 7-1). MAD data were collected using the inverse beam mode. Data were reduced using the programs DENZO and SCALEPACK⁴², then processed using the CCP4 suite⁴³.

Phase determination and structure refinement. We used a combination of multiwavelength anomalous dispersion (MAD)⁴⁴ from selenomethionyl residues (SeMet) incorporated into the NFAT and multiple isomorphous replacement (MIR) from iodinated DNA bases and a conventionally prepared methylmercury derivative. Native I was isomorphous to all of the derivatives, and this data set was used for initial phase determination and model building. Mercury positions (four sites) were found by inspecting the anomalous difference Patterson map. Subsequently, iodine and selenium positions were determined by the difference Fourier method. Refinement of heavy-atom parameters in MIR and MAD and phase calculation were performed using MLPHARE⁴⁵. MAD and MIR phases were combined using SIGMAA⁴³. The phase-combined map was further improved by solvent flattening using DM⁴³.

The overall figure of merit was 0.543 at 20–3.1 Å resolution for the initial map. About 85% of the total model was built into this map. DNA was built by fitting a standard B DNA to the density with appropriate adjustments. Fos and Jun were built by matching the previously determined model with the map as rigid bodies. Whereas Jun could be fit as one piece, the basic region and leucine zipper of Fos had to be fit as two separate pieces. NFAT was built directly in O by the baton-build utility⁴⁵. Iterative rebuilding and refinement (positional and simulated annealing) were performed using the program O and XPLOR⁴⁶. A partially refined model in the native I lattice was transformed to other lattices (native II and native III) by one round of rigid body refinement in XPLOR. Maps calculated in the three different lattices were averaged by the program MAVER⁴⁷, and the model was further adjusted using the triple-averaged map. Phase and refinement statistics are given in Table 1. Final models have very good geometry as examined by PROCHECK⁴³ (all residues have ϕ/ψ angles in the 'allowed regions' of a Ramachandran plot, with 86% of them in the 'most

favoured region'). Simulated-annealing omit maps have been computed to check several regions of the model, including the E'F loop, the CX loop and the AB loop.

Received 17 September; accepted 23 December 1997.

- Weiss, A. & Littman, D. R. Signal transduction by lymphocyte antigen receptors. *Cell* **76**, 263–274 (1994).
- Crabtree, G. R. & Clipstone, N. A. Signal transmission between the plasma membrane and nucleus of T lymphocytes. *Annu. Rev. Biochem.* **63**, 1045–1083 (1994).
- Jain, J., Loh, C. & Rao, A. Transcriptional regulation of the IL-2 gene. *Curr. Opin. Immunol.* **7**, 333–342 (1995).
- Serfling, E., Avots, A. & Neumann, M. The architecture of the interleukin-2 promoter: a reflection of T lymphocyte activation. *Biochim. Biophys. Acta* **1263**, 181–200 (1995).
- Shaw, J. P. et al. Identification of a putative regulator of early T cell activation genes. *Science* **241**, 202–205 (1988).
- Rao, A., Luo, C. & Hogan, P. G. Transcription factors of the NFAT family: regulation and function. *Annu. Rev. Immunol.* **15**, 707–747 (1997).
- Schreiber, S. L. & Crabtree, G. R. The mechanism of action of cyclosporin A and FK506. *Immunol. Today* **13**, 136–142 (1992).
- McCaffrey, P. G. et al. Isolation of the cyclosporin-sensitive T cell transcription factor NFATp. *Science* **262**, 750–754 (1993).
- Northrop, J. P. et al. NF-AT components define a family of transcription factors targeted in T-cell activation. *Nature* **369**, 497–502 (1994).
- Masuda, E. S. et al. NFATx, a novel member of the nuclear factor of activated T cells family that is expressed predominantly in the thymus. *Mol. Cell Biol.* **15**, 2697–2706 (1995).
- Hoey, T., Sun, Y. L., Williamson, K. & Xu, X. Isolation of two new members of the NF-AT gene family and functional characterization of the NF-AT proteins. *Immunity* **2**, 461–472 (1995).
- Jain, J., Burgeon, E., Badalian, T. M., Hogan, P. G. & Rao, A. A similar DNA-binding motif in NFAT family proteins and the Rel homology region. *J. Biol. Chem.* **270**, 4138–4145 (1995).
- Chytil, M. & Verdine, G. L. The Rel family of eukaryotic transcription factors. *Curr. Opin. Struct. Biol.* **6**, 91–100 (1996).
- Shaw, K. T. et al. Immunosuppressive drugs prevent a rapid dephosphorylation of transcription factor NFAT1 in stimulated immune cells. *Proc. Natl Acad. Sci. USA* **92**, 11205–11209 (1995).
- Timmerman, L. A., Clipstone, N. A., Ho, S. N., Northrop, J. P. & Crabtree, G. R. Rapid shuttling of NF-AT in discrimination of Ca²⁺ signals and immunosuppression. *Nature* **383**, 837–840 (1996).
- Luo, C. et al. Interaction of calcineurin with a domain of the transcription factor NFAT1 that controls nuclear import. *Proc. Natl Acad. Sci. USA* **93**, 8907–8912 (1996).
- Shibasaki, F., Price, E. R., Milan, D. & McKeon, F. Role of kinases and the phosphatase calcineurin in the nuclear shuttling of transcription factor NF-AT4. *Nature* **382**, 370–373 (1996).
- Beals, C. R., Clipstone, N. A., Ho, S. N. & Crabtree, G. R. Nuclear localization of NF-ATc by a calcineurin-dependent, cyclosporin-sensitive intramolecular interaction. *Genes Dev.* **11**, 824–834 (1997).
- Liu, J. et al. Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell* **66**, 807–815 (1991).
- Jain, J., McCaffrey, P. G., Valge-Archer, V. E. & Rao, A. Nuclear factor of activated T cells contains Fos and Jun. *Nature* **356**, 801–804 (1992).
- Jain, J., Miner, Z. & Rao, A. Analysis of the preexisting and nuclear forms of nuclear factor of activated T cells. *J. Immunol.* **151**, 837–848 (1993).
- Cockrell, P. N. et al. Human granulocyte-macrophage colony-stimulating factor enhancer function is associated with cooperative interactions between AP-1 and NFATp/c. *Mol. Cell Biol.* **15**, 2071–2079 (1995).
- Jain, J. et al. The T-cell transcription factor NFATp is a substrate for calcineurin and interacts with Fos and Jun. *Nature* **365**, 352–355 (1993).
- Muller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L. & Harrison, S. C. Structure of the NF- κ B p50 homodimer bound to DNA. *Nature* **373**, 311–317 (1995).
- Ghosh, G., van Duyne, G., Ghosh, S. & Sigler, P. B. Structure of NF- κ B p50 homodimer bound to a κ B site. *Nature* **373**, 303–310 (1995).
- Wolfe, S. A. et al. Unusual Rel-like architecture in the DNA-binding domain of the transcription factor NFATc. *Nature* **385**, 172–176 (1997).
- Chen, L. et al. Only one of the two DNA-bound orientations of AP-1 found in solution cooperates with NFATp. *Curr. Biol.* **5**, 882–889 (1995).
- Bork, P., Holm, L. & Sander, C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309–320 (1994).
- Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: crystal structure of the protein-DNA complex. *Cell* **71**, 1223–1237 (1992).
- König, P. & Richmond, T. J. The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.* **233**, 139–154 (1993).
- Glover, J. N. & Harrison, S. C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* **373**, 257–261 (1995).
- Seeman, N. C., Rosenberg, J. M. & Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA* **73**, 804–808 (1976).
- Pabo, C. O. & Sauer, R. T. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053–1095 (1992).
- Chen, L. *Functional and Structural Studies of DNA Binding Proteins*. Thesis, Harvard Univ., (1994).
- Sun, L. J., Peterson, B. R. & Verdine, G. L. Dual role of the nuclear factor of activated T cells insert region in DNA recognition and cooperative contacts to activator protein 1. *Proc. Natl Acad. Sci. USA* **94**, 4919–4924 (1997).
- Peterson, B. R., Sun, L. J. & Verdine, G. L. A critical arginine residue mediates cooperativity in the contact interface between transcription factors NFAT and AP-1. *Proc. Natl Acad. Sci. USA* **93**, 13671–13676 (1996).
- Yamamoto, K. R., Pearce, D., Thomas, J. & Miner, J. N. in *Transcriptional Regulation* (eds McKnight, S. L. & Yamamoto, K. R.) 1169–1192 (Cold Spring Harbor Laboratory Press, NY, 1992).
- Johnson, A. in *Transcriptional Regulation* (eds McKnight, S. L. & Yamamoto, K. R.) 975–1006 (Cold Spring Harbor Laboratory Press, NY, 1992).
- Tjian, R. & Maniatis, T. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* **77**, 5–8 (1994).
- Thanos, D. & Maniatis, T. Virus induction of human IFN β gene expression requires the assembly of an enhancosome. *Cell* **83**, 1091–1100 (1995).
- Garrity, P. A., Chen, D., Rothenberg, E. V. & Wold, B. J. Interleukin-2 transcription is regulated in vivo at the level of coordinated binding of both constitutive and regulated factors. *Mol. Cell Biol.* **14**, 2159–2169 (1994).
- Otwiniński, Z. in *Proceedings of the CCP4 study weekend* (eds Sawyer, L., Isaacs, N. & Burley, S.) 56–62 (SERC Daresbury Laboratory, Warrington, 1993).
- Collaborative Computational Project Number 4. The CCP suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–776 (1994).
- Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
- Jones, T. A., Bergdoll, M. & Kjeldgaard, M. in *Crystallographic Computing and Modeling Methods in Molecular Design* (eds Bugg, C. & Ealick, S.) (Springer, New York, 1989).
- Brunger, A. T. *X-PLOR Version 3.0: A System for Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
- Kleywegt, G. & Jones, T. A. in *From First Map to Final Model* (eds Bailey, S., Hubbard, R. & Walker, D.) (Daresbury Laboratory, Warrington, 1994).
- Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
- Carson, M. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).

Acknowledgements. We thank T. Hsieh, B. Willard and N. Sinitskaya for help with protein and DNA preparation; K. Leong, T. Stehle, J. Wang, R. Chopra and M. Jacobs for help with data collection and computation; J. Jain, G. Verdine, S. Wolfe, D. Erlanson, P. Zhou, L. Sun, B. Peterson and C. Vaughan for discussions; and L. Berman, Z. Yin and M. Soltis for synchrotron technical assistance. L.C. was supported by the Cancer Research Fund of the Damon Runyon-Walter Winchell Foundation Fellowship, and by The Medical Foundation. S.C.H. is an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to S.C.H. or L.C. (e-mail: schadmin@crystal.harvard.edu). Coordinates have been deposited at the Brookhaven Protein Data Bank, accession code 1a02, and are also available by e-mail from lchen@xtal1200.harvard.edu.

Two distinct populations of Kuiper-belt objects

S. C. Tegler* & W. Romanishin†

* Department of Physics & Astronomy, Northern Arizona University, Flagstaff, Arizona 86011, USA

† Department of Physics & Astronomy, University of Oklahoma, Norman, Oklahoma 73019, USA

The discovery of the first member of the Kuiper belt¹—a formerly hypothetical ancient reservoir of objects located beyond Neptune's orbit—started a revolution in our understanding of the outer Solar System: there is no longer a sharp edge at Pluto's orbit. About 60 Kuiper-belt objects, intermediate in size between comets and planets, are now known² to exist on stable circular orbits around the Sun, and no doubt many more objects await discovery. But owing to the recent discovery and intrinsic faintness of these objects, little has been done to explore their physical and chemical properties^{3–8}. Here we report the results of a two-year survey of the broad-band optical colours of about one-quarter of the known Kuiper-belt objects. We find that their colours indicate the presence of two distinct populations: one consists of objects whose surface colours are only slightly redder than the colour of the Sun, while the other consists of the reddest objects known in the Solar System.

Our survey uses Harris B (450 nm), V (550 nm) and R (650 nm) 5-inch $\times$ 5-inch glass filters in front of a 1,200 $\times$ 800 pixel CCD (charge coupled device) camera at the $f/9$ Cassegrain focus of the Steward Observatory 2.3-m telescope. The CCD has outstanding quantum efficiency, particularly in the blue where it exceeds 95%. This high quantum efficiency greatly aided our blue photometry of these faint objects. We used the CCD in a 2 $\times$ 2 binning mode, yielding 600 $\times$ 400 pixel images, covering 3 $\times$ 2 arcmin on the sky at 0.30 arcsec per pixel. Here we will use "pixel" to refer to the binned 0.3-arcsec pixels. The typical full-width at half-maximum of the stellar point-spread function is 1.5 arcsec.

Observations of Kuiper-belt objects (KBOs) present a formidable challenge for a photometrist. Their faintness requires us to image small areas of the sky to faint brightness levels; at such levels there are numerous faint stars and galaxies in the images. The Earth's revolution about the Sun results in the motion of KBOs relative to the "fixed" background of stars and galaxies in the images (~ 3 arcsec h^{-1}). Hence, KBOs sometimes pass near background stars and galaxies, and the stellar and galactic light sometimes contaminates the light from KBOs. Finally, some KBOs show brightness variations that are probably the result of the rotation and irregular shape of the KBOs or of albedo variations over their surfaces. Observations of these faint, moving, variable objects require great care.

We have developed an observational and image processing procedure customized for photometry of KBOs^{5–7}. The faintness and motion of these objects pushes the optimum exposure time in opposite directions. On the one hand, the faintness of the objects pushes us towards exposure times that are as long as possible, so that we can maximize the signal-to-noise ratio of the data. On the other hand, the motion of KBOs pushes us towards short exposure times, otherwise the light from a KBO is spread over many pixels, thereby degrading the signal-to-noise ratio. We have found that taking individual 300-s exposures works best. Specifically, we obtain between two and 20 images per object, per filter, per night, depending on the brightness of the object, and then compute the mean and the uncertainty in the mean of the instrumental magnitudes.

With two exceptions (the bright objects 5145 Pholus and 1997

CU26), all of the objects in our survey have been observed on 2–4 nights. For the fainter objects, we averaged a small number (typically three) of 300-s images; taken together in time, this yields an effective exposure time of typically 900 s. Because of the motion of KBOs, two combined images were created for each group. One group had the individual images registered on the stars, and the other group had the images registered on the KBO before combining (averaging) the images. By combining the images in this way, the motion of the KBO from image to image did not contribute to any smearing of the KBO image in the combined image.

To maximize the signal-to-noise ratio of the measurements in the 300- or 900-s images, we have applied a well known technique of stellar photometry, as follows. KBOs are so faint that the uncertainty in their instrumental magnitudes is dominated by sky noise. Under such conditions, it is common in stellar photometry to use a rather small aperture when making the instrumental magnitude measurement, in order to minimize the sky noise; a correction is then made to the 'total' instrumental magnitude using observations in small and large apertures of a brighter comparison star in the same field. We use this method to get instrumental magnitudes for the KBOs. For each 300-s or 900-s image, we measured the KBO with a small aperture (typically 1.8 arcsec diameter; 3 pixel radius); we also measured a brighter comparison star with the same aperture, and again with a larger aperture (typically 9.0 arcsec diameter; 15 pixel radius) that contained all the light from the comparison star. We then applied the magnitude correction from the small and large apertures of the comparison-star measurement to the KBO measurement, yielding the 'total' instrumental magnitude for the KBO. For averaged 900-s images, the magnitude corrections were

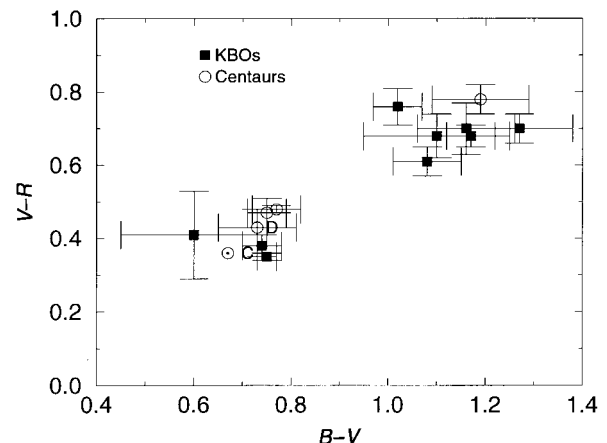


Figure 1 Colour-colour plot of the KBOs (filled squares) and Centaurs (open circles) in our survey. The colour of the Sun is represented by the standard solar symbol ($\odot$) at $B - V = 0.67$ and $V - R = 0.36$. The average colours of C and D class asteroids¹⁶ are represented by bold letters. KBOs and Centaurs fall into two discrete populations: a solar to slightly red population similar to C and D class asteroids, and an extraordinary red population. We note that the $V - R$ colours for three additional objects, 1996 RQ20, 1994 JR1 and 7066 Nessus, fall into the two populations; see Table 1. The error bars are the uncertainty in the mean, $\sigma/\sqrt{n}$, where σ is the standard deviation of n 300-s or 900-s exposures. Photon noise, sky noise and "unseen" background objects just below our detection limit are sources of random error that contribute to the size of the error bars. We estimate that we can detect background objects as faint as 5–10% the brightness of a KBO or Centaur (for most of the objects in our survey). Therefore, "unseen" background objects should affect our photometry by less than 5–10% (0.05–0.10 mag). Unseen background objects should affect the colours by far less, because the objects affect the B, V and R filters in the same sense, and to tend to cancel out in the colours. As the scatter in our measurements is ≤ 0.05 mag for most objects, we conclude that the effect of "unseen" background objects on our measurements is negligible (< 0.05 mag).

determined from the images registered on the stars, and the KBO measurements were made from the images registered on the KBOs. The aperture correction is highly dependent on the seeing. Over two years we have been subject to a variety of seeing conditions and hence the aperture corrections range between ~ 0.5 and 1.0 mag. Typically, the correction is ~ 0.7 mag. A comparison of several bright field stars on a given image indicates that the uncertainty in the aperture correction for an image is ≤ 0.01 mag. Aperture photometry was done with the PHOT task in the IRAF DIGIPHOT package¹⁷. For all photometry, the sky value was found in an annulus from 20 to 30 pixel radius. The sky value was the peak of a gaussian fit to the histogram of pixel values in the sky annulus. If necessary, the sky annulus was first 'cleaned' of objects which might bias the sky measurement by replacing them with a patch of nearby sky. Because we typically spent many hours per night looking at a single object, we are able to examine the paths of the objects for faint galaxies and stars that would corrupt our photometry. We discarded any KBO images that passed in front of faint background stars or galaxies.

The KBO instrumental magnitudes were corrected to zero air-mass using nightly extinction coefficients. We observed three or four standard stars, two to three times per night, over a range of ~ 1 airmass for the derivation of our extinction coefficients. Our coefficients are consistent with standard values for Kitt Peak. Finally, KBO instrumental magnitudes were placed on the Kron-Cousins system using transformation equations derived from observations of standard stars in the M67, PG0231 + 051, PG1633 + 099 and PG2213 - 006 fields^{9,10}. As a check on our colour transformation equations and the V zero-point equation, we treated a few additional standard stars as objects and derived V magnitudes and colours for comparison with published values. The comparison suggests our equations are accurate to the 0.01 mag level.

The results of our survey are unexpected; they are summarized in Table 1 and in the colour-colour plot of Fig. 1. The reddest objects are towards the upper right-hand corner of the figure, and the bluest objects are towards the lower left-hand corner. Colours for KBOs and Centaurs (recent refugees from the Kuiper belt¹¹ on outer planet crossing orbits) are represented by filled squares and open circles, respectively. Clearly, there is evidence for two discrete colour populations. The population in the upper-right corner of Fig. 1 is extraordinarily red; the bright Centaur 5145 Pholus is a member of this population. The population in the lower-left corner of Fig. 1 contains objects with neutral (solar) colours to slightly red colours (similar to C, P and D class asteroids) and includes the bright Centaur 2060 Chiron. We point out that three objects (1996 RQ20, 1994 JR1 and 7066 Nessus) in Table 1 do not appear in Fig. 1, which is because we have not yet measured their $B - V$ colours. We note the $V - R$ colours of these three objects fall in the two populations. The Sun is plotted for reference and is represented by the standard solar symbol ($\odot$) at $B - V = 0.67$ and $V - R = 0.36$.

Table 1 Colours of KBOs and Centaurs

Object	Class	$V - R$	$\sigma/\sqrt{h}$	$B - V$	$\sigma/\sqrt{h}$	Ref.
1997 CS29	KBO	0.61	0.04	1.08	0.07	
1996 TS66	KBO	0.76	0.05	1.02	0.05	
1996 TQ66	KBO	0.70	0.07	1.16	0.10	
1996 TP66	KBO	0.68	0.03	1.17	0.05	
1996 TO66	KBO	0.38	0.03	0.74	0.04	
1996 TL66	KBO	0.35	0.01	0.75	0.02	
1996 RQ20	KBO	0.44	0.05			
1995 HM5	KBO	0.41	0.12	0.60	0.15	
1994 TB	KBO	0.68	0.06	1.10	0.15	6
1994 JR1	KBO	0.36	0.08			
1993 SC	KBO	0.70	0.04	1.27	0.11	6
1997 CU26	Cen.	0.48	0.01	0.77	0.05	
1995 GO	Cen.	0.47	0.04	0.75	0.04	5
1995 DW2	Cen.	0.43	0.05	0.73	0.08	
7066 Nessus	Cen.	0.77	0.05			5
5145 Pholus	Cen.	0.78	0.04	1.19	0.10	5

The only other colour survey with as many objects as ours shows a diversity of colours³, rather than two colour populations. We believe that the difference is due to the fact that our photometry is significantly more accurate. Our uncertainties in both the $B - V$ and $V - R$ measurements are less than, or equal to, 0.05 mag for most KBOs (see Table 1); in contrast, the uncertainties in the other colour survey³ are typically between 0.1 and 0.2 mag for most KBO measurements in that sample (see Table 5 in ref. 3). As the populations in Fig. 1 are separated by 0.3 – 0.4 mag, uncertainties of 0.05 mag are essential to detect the discrete populations. The accuracy of our measurements is certainly due to the fact that (1) we obtain on average a dozen 300-s images per filter, per object, and (2) we measure the brightnesses of KBOs using the central, highest signal-to-noise pixels in the KBO images, and scale to the total brightness of the KBOs using bright comparison stars in the same fields.

What is the mechanism responsible for the bimodal colour distribution? The lack of any correlation between the colours of the objects in Table 1 and their inclination angle, eccentricity, perihelion distance, semi-major axis, or absolute magnitude, makes it difficult to point to a specific mechanism in a conclusive way. What is clear is that the surface evolution of KBOs is a complex problem. Laboratory simulations show that the impact of solar ultraviolet photons, solar-wind and solar-flare particles, and cosmic rays on icy surfaces in the outer Solar System alters the chemical composition of those surfaces. For example, data from laboratory simulations indicate that bright, hydrocarbon-rich ices initially redden with radiation exposure, but then become grey with increasing radiation dose^{12,13}. In addition, the impact of particles ranging in size from micrometeorites to kilometre-sized bodies results in cratering which further alters the chemical composition of the surfaces.

It is also possible that coma formation may modify the surfaces of Centaurs and KBOs. For example, the Centaurs 2060 Chiron and 5145 Pholus have nearly identical perihelion distances, but the latter has an extraordinarily red colour and shows no evidence of coma activity, and the former has a grey (solar) colour and does show coma activity¹⁴. The coma activity of 2060 Chiron is not correlated with perihelion distance¹⁵. Hence, something other than solar heating alone is driving the coma formation of 2060 Chiron; perhaps other Centaurs and KBOs will be found to have coma activity. Further observational, theoretical and laboratory work is necessary to explain conclusively the unexpected bimodal colour distribution of KBO and Centaur surfaces. $\square$

Received 2 December 1997; accepted 20 January 1998.

- Jewitt, D. & Luu, J. Discovery of the candidate Kuiper belt object 1992 QB1. *Nature* **362**, 730–732 (1993).
- Marsden, B. G. *List of Transneptunian Objects* (1997); available at <http://cfa-www.harvard.edu/cfa/ps/lists/TNOs.html>
- Luu, J. & Jewitt, D. Color diversity among the centaurs and kuiper belt objects. *Astron. J.* **112**, 2310–2318 (1996).
- Brown, W. & Luu, J. X. CCD photometry of the centaur 1995 GO. *Icarus* **126**, 218–224 (1997).
- Romanishin, W., Tegler, S. C., Levine, J. & Butler, N. BVR photometry of centaur objects 1995 GO, 1993 HA2, and 5145 Pholus. *Astron. J.* **113**, 1893–1898 (1997).
- Tegler, S. C. & Romanishin, W. The extraordinary colors of trans-neptunian objects 1994 TB and 1993 SC. *Icarus* **126**, 212–217 (1997).
- Tegler, S. C. et al. Photometry of the trans-neptunian object 1993 SC. *Astron. J.* **114**, 1230–1233 (1997).
- Weintraub, D. A., Tegler, S. C. & Romanishin, W. Visible and near infrared photometry of the centaur objects 1995 GO and 5145 Pholus. *Icarus* **128**, 456–463 (1997).
- Landolt, A. U. UBVR photometric standard stars in the magnitude range $11.5 < V < 16.0$ around the celestial equator. *Astron. J.* **104**, 340–371 (1992).
- Richmond, M. W. Automated photometry at Leuschner observatory. *Int. Amateur-Professional Photoelectr. Photom. Commun.* **55**, 21–30 (1994).
- Stern, A. & Campins, H. Chiron and the centaurs: escapees from the Kuiper belt. *Nature* **382**, 507–510 (1996).
- Andronico, G., Baratta, G. A., Spinella, F. & Strazzulla, G. Optical evolution of laboratory produced organics: applications to Phoebe, Iapetus, outer belt asteroids, and cometary nuclei. *Astron. Astrophys.* **184**, 333–336 (1987).
- Thompson, R. W., Murray, B. G. J. P. T., Khare, B. N. & Sagan, C. Coloration and darkening of methane clathrate and other ices by charged particle irradiation: applications to the outer solar system. *J. Geophys. Res.* **92**, 14933–14947 (1987).
- Bus, S. J., Schleicher, B. W., E. & A'Hearn, M. F. Detection of CN emission from 2060 Chiron. *Science* **251**, 774–777 (1991).

15. Bus, S. J., A'Hearn, M. F., Bowell, E. & Stern, S. A. 2060 Chiron: evidence for activity near aphelion. *Icarus* (submitted).
16. Degewij, J. & Van Houten, C. J. in *Asteroids* (ed. Gehrels, T.) 417–435 (Univ. Arizona Press, Tucson, 1979).
17. Tody, D. in *Astronomical Data Analysis Software and Systems II* (eds Hanisch, R. J., Brissenden, R. J. V. & Barnes, J.) 173 (ASP Conf. Ser. 52, Tucson, 1993).

Acknowledgements. We thank the NASA Origins of Solar Systems program for support of this work, and the Steward Observatory Telescope Allocation Committee for consistent allocation of telescope time.

Correspondence and requests for materials should be addressed to S.C.T. (e-mail address: tegler@proton.phy.nau.edu).

Subsurface charge accumulation imaging of a quantum Hall liquid

S. H. Tessmer^{*†}, P. I. Glicofridis^{*}, R. C. Ashoori^{*}, L. S. Levitov^{*} & M. R. Melloch[‡]

^{*} Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[‡] Department of Electrical Engineering, Purdue University, West Lafayette, Indiana 37907, USA

The unusual properties of two-dimensional electron systems that give rise to the quantum Hall effect have prompted the development of new microscopic models for electrical conduction^{1–6}. The bulk properties of the quantum Hall effect have also been studied experimentally using a variety of probes including transport^{7,8}, photoluminescence^{9,10}, magnetization¹¹ and capacitance^{12,13} measurements. However, the fact that two-dimensional electron systems typically exist some distance (about 1,000 Å) beneath the surface of the host semiconductor has presented an important obstacle to more direct measurements of microscopic electronic structure in the quantum Hall regime. Here we introduce a cryogenic scanning-probe technique—‘subsurface charge accumulation’ imaging—that permits very high resolution examination of systems of mobile electrons inside materials. We use this technique to image directly the nanometre-scale electronic structures that exist in the quantum Hall regime.

The subsurface charge accumulation (SCA) probe measures the local accumulation of charge in a two-dimensional electron system (2DES) in response to an applied a.c. excitation. We use it to examine the quantum Hall system over areas as large as several micrometres, and with a spatial resolution of ~900 Å. For magnetic fields near Hall plateaux, fine structures appear in the SCA images. These structures move and contort rapidly as the field is varied, even though they reappear on successive Hall plateaux.

In contrast to other techniques used to study a 2DES located beneath an insulating surface^{14,15}, SCA imaging is an a.c. method^{16–19} and has significant advantages for the study of the quantum Hall effect over measurements which image the distribution of static charges. As samples contain static charges from both the surface and dopant layers underneath the surface, these regions can mask charging features in the two-dimensional layer¹⁵. SCA imaging, operating at frequencies of ~100 kHz (very low compared to other a.c. methods), is uniquely suited for studying the accumulation of mobile charges in the 2DES in magnetic fields.

SCA microscopy of the 2DES proceeds as follows. A sharp conducting tip is positioned ~50 Å above the sample surface and connected to a highly sensitive charge detector (Fig. 1). An a.c. excitation voltage is applied to the 2DES through an ohmic contact (2 mm away from the tip), and no other contacts are made to the sample. The excitation causes charge to flow in and out of the 2DES,

in turn inducing charge to flow in and out of the tip. In scanning the tip and using synchronous lock-in detection of the induced charge, we obtain two images, one for charge accumulating in-phase (Q_{in}) and the other 90° (lagging) out-of-phase (Q_{out}) with the excitation. The microscope also operates in other modes allowing determination of surface topography (tunnelling mode) and local accumulations of static charge (Kelvin probe mode)²⁰.

As a test of our microscope for studying quantum Hall features, we first use it to locally perturb the 2DES. If large voltages (>2 V) are applied to the tip, static and immobile charge can be deposited on the exposed surface of the sample or in the donor layer between the surface and the 2DES, thereby locally altering the electron density in the 2DES. Figure 2 shows a series of six Q_{in} (in-phase charging) SCA images centred on such a location. In these images, dark and bright represent respectively small and large Q_{in} signals. We note the dark spot which appears as the field is raised to 1.5 T and grows for larger fields.

The origin of the dark spot can be understood by considering basic properties of the 2DES in the quantum Hall state. Any region containing a density which precisely fills an integer number of Landau levels is ‘incompressible’ and does not accumulate charge or conduct electricity. Such regions therefore appear dark in our images. At the centre of the images shown in Fig. 2, the electron density has a broad minimum (with the density reduced to about one-quarter of the bulk value). At the borders of the dark spot, the density is such that one Landau level is filled, preventing charge transport to the interior. Outside the dark spot, the density of electrons is high enough so that more than one level is still filled, and charge accumulates freely.

Incompressible regions such as this dark spot appear in some theories of the quantum Hall effect^{5,21,22}; they are introduced to explain the transitions between the quantum Hall plateaux. At

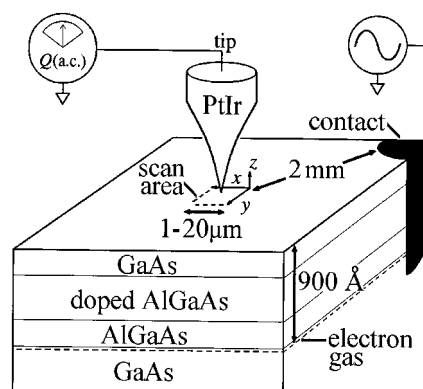
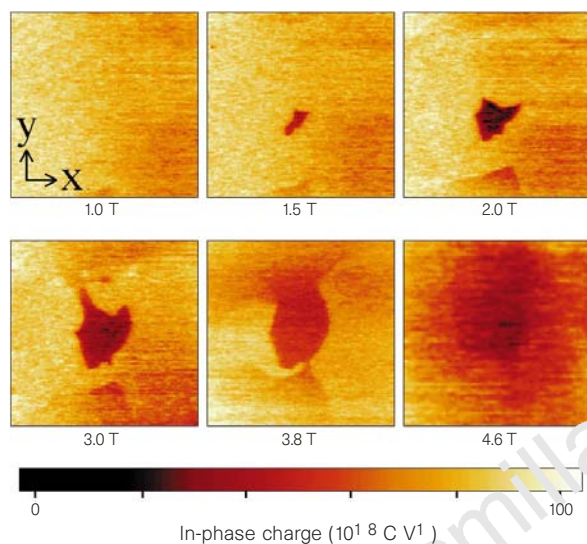


Figure 1 Diagram of the sample and measurement configuration. The two-dimensional electron system (2DES) exists at the interface of $Al_{0.3}Ga_{0.7}As$ and GaAs 900 Å beneath the surface. The a.c. excitation applied to the 2DES causes this layer to charge and discharge due to the self-capacitance of the layer. A sharp metal tip is connected to a high electron mobility transistor (HEMT) with extremely high sensitivity to electrical charge³⁰ (0.01 electrons $\sqrt{Hz}$) and scanned 50 Å above the sample surface. SCA microscopy produces an image of charge accumulating in the 2DES by detecting charge induced on the tip due to charging in the layer. The spatial resolution is given approximately by the depth of the 2DES below the sample surface. The microscope operates with the sample immersed in liquid 3He ($T = 300$ mK) and has a 20-μm-long scan range. The static electron density in the 2DES is $3.0 \times 10^{11} cm^{-2}$ and the transport mobility is $450,000 cm^2 V^{-1} s^{-1}$. The small amplitude of the 100 kHz excitation applied to the 2DEG (1–4 mV) and the smallness of the capacitance between the tip and the 2DES (compared to the self-capacitance of the 2DES) ensure that the tip produces practically no perturbation of the 2D system.

[†] Present address: Michigan State University, East Lansing, Michigan 48824, USA.

magnetic fields just below the quantum Hall plateaux, the theory predicts that the 2DES contains a system of incompressible droplets. As in the case of Fig. 2, as the field is increased, these droplets grow and begin to merge. Thus, entering the Hall plateau, the 2DES undergoes a percolation transition with the network of merged droplets impeding conduction across the sample, and thus producing a quantum Hall plateau^{1,4}. However, in the absence of long-range potential fluctuations, theories do not require the existence of droplets at fixed positions^{23–25}. The local charging observed in Fig. 2 displays sharp contrast, consistent with the concept of distinct insulating and conducting regions.



Having shown that the microscope can resolve quantum Hall features, we now use it to examine an unperturbed system. In contrast to the perturbed region, the SCA images only develop features over narrow ranges in magnetic field around integer Landau level filling (ν) values. As the magnetic field is adjusted towards the centre of the Hall plateau, the charge accumulation seems to diminish roughly uniformly at all locations and detailed spatial patterns of charging appear.

Here we concentrate on the $\nu = 4$ Hall plateau which is centred around 3.04 T in this sample. Images taken at fields of more than 0.15 T away from the centre of the Hall plateau show no features.

Figure 2 In-phase (Q_{in}) SCA images of a region of perturbed electron density in the 2DES. Images ($2.7 \times 2.7 \mu\text{m}$) are shown for various magnetic fields applied perpendicular to the 2D plane. Static charge measurements made in Kelvin probe mode show the 2D electron density has a shallow minimum at the centre of the images and increases to the bulk value away from the centre. For a magnetic field as high as 1.0 T, the SCA image has few features. But at 1.5 T, the signal drops to a lower value when the tip is located near the centre, displaying a small feature (dark spot) which grows with increasing magnetic field. The density reduction at the spot centre can be explained by considering the magnetic field at which the black spot appears and by following its subsequent growth with field. In the unperturbed 2DES, the filling factor $\nu = 2$ (one Landau level filled with both spin-up and spin-down electrons) occurs at 6.1 T. The black spot appears at 1.5 T, and grows monotonically as the field is raised. As this growth is observed to occur beyond the field for $\nu = 3$ in the bulk, we conclude that the spot must first appear when the filling factor at the centre of the perturbed region is $\nu = 2$. The spot emerges at one-quarter of the field required for $\nu = 2$ in the bulk, suggesting that the density at the spot has been reduced to about one-quarter of the bulk density.

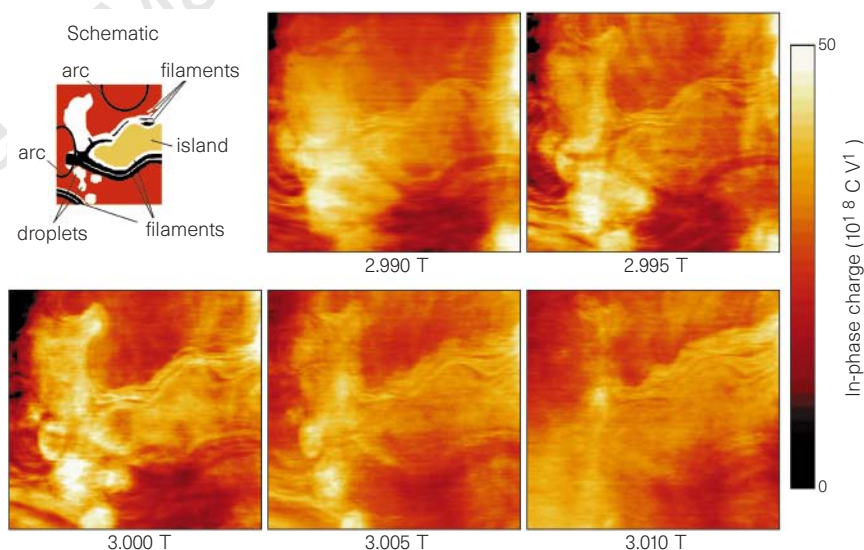


Figure 3 Sequence of five in-phase SCA images for an unperturbed region of the 2DES at $\nu \approx 4.05$. The images ($3.3 \times 3.3 \mu\text{m}$) are for magnetic fields differing by 0.005 T, and each image was obtained using 5 h of signal averaging. Kelvin probe measurements were used to find a region unperturbed by static charge fluctuations. The contact potential was measured with the microscope in Kelvin probe mode, and the tip bias (+0.80 V) was then set to eliminate d.c. fields between the tip and the 2DES. To enhance our ability to display small-scale features, a background plane has been subtracted from each image. This does not alter the images significantly nor create structural artefacts in the images. We note that, despite the great sensitivity of the structures to changes in magnetic field, the structures do not change at all with time. The expected spatial resolution of

the probe is 900 Å, consistent with the smallest features displayed in the images. 'Filaments' (see schematic diagram at top left) of high or low charge accumulation appear in each of the images. In the 3.000 T image, these filaments are separated by $\sim 1,500$ Å and meander in unison for a distance of at least $2 \mu\text{m}$ in the image. The filaments exist at the boundary between broader regions of high and low charge accumulation. Some filaments appear as broad dark bands that wind all the way across the image. Also, a grouping of high charge accumulating bulbous 'droplets' is clearly shown in the images at 3.000 T and 3.005 T. The droplets shrink rapidly in size as the field strength is increased. We also note that 'arcs' or 'loops' appear in the images. The upper arc shown in the schematic exists in the image at 2.995 T, and it appears to follow the circumference of a loop of diameter $\sim 1.5 \mu\text{m}$.

This indicates that spatial fluctuations in the charging (per volt of excitation amplitude) are less than $1 \times 10^{-18} \text{ C V}^{-1}$. For comparison, the dark region observed at 3.20 T in the perturbed area (Fig. 2) displayed a reduction of $50 \times 10^{-18} \text{ C V}^{-1}$. In contrast, clear and intricate charging features of amplitudes in the range of $20 \times 10^{-18} \text{ C V}^{-1}$ are observed for fields near the plateau centre. The features are extraordinarily sensitive to magnetic field, with little correlation between the images at fields differing by 0.020 T (or 0.7%).

In Fig. 3 we show a series of five images taken at magnetic fields spaced by 0.005 T. The images reveal a startling amount of detail (see Fig. 3 legend). One may quickly ascertain that the features all indeed arise from charging structure within the two-dimensional layer resulting from the quantum Hall effect. There is no model for electronic structure in the donor layer or on the sample surface having such profound field sensitivity. Moreover, the structures are only seen when the magnetic field is near Hall plateaux. Last, features redevelop when the magnetic field is increased by precisely a factor of two, corresponding to the change from $\nu = 4$ to $\nu = 2$; the path followed by the filaments at 3.000 T (see schematic in Fig. 3) is once again followed by filaments only at precisely 6.000 T (near $\nu = 2$). No known physical process would give rise to such periodicity other than the quantum Hall effect.

What is the origin of the contrast in the 2DES? The observed structure may arise from spatial variation in compressibility of the 2DES (that is, some regions of the 2DES can absorb more charge than others), or may be due to differences in local charging rates. It is possible to distinguish between these two mechanisms by comparing the contrast in the Q_{in} and Q_{out} images.

A simple approach to the 2DES is to consider it as an RC system (where R is the resistance of the layer, and C is its self-capacitance).

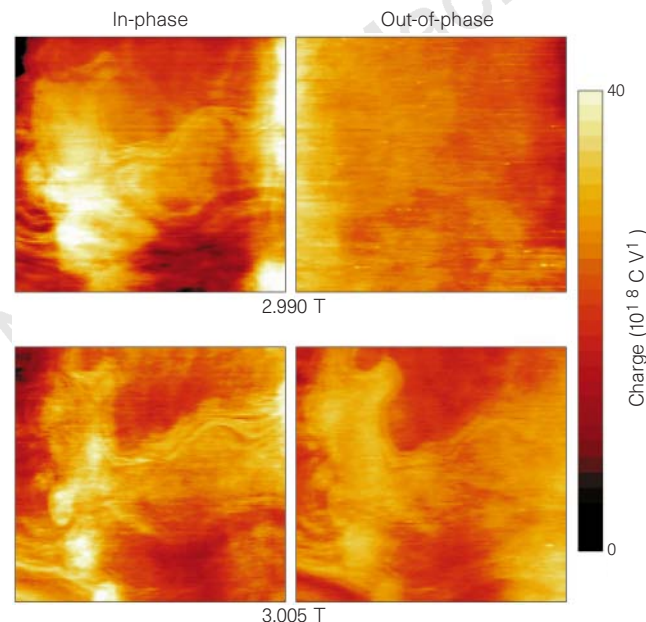


Figure 4 Q_{in} and Q_{out} images at 2.990 T (top row) and 3.005 T (bottom row). We note that Q_{out} is nearly featureless at 2.990 T while Q_{in} displays detailed structure. Increasing the field by only 150 gauss to 3.005 T dramatically alters the phase of the image. The overall resistance of the sample increases as the field is raised from 2.990 to 3.005 T, and the model for charging of the 2DES (see text) explains the observed phase shift. At 3.005 T, the contrast seen in Q_{out} always has the same sign as that in Q_{in} . Based on a simple RC charging model, it is possible to show that this observation is consistent with our conclusion that the structures in the images arise from variations in local compressibility and not from variation in local charging rates.

In our experiment, the resistivity of the sample varies as a function of magnetic field, and becomes extremely large at integer filling. At these fields, the roll-off frequency $f_0 = 1/(2\pi RC)$ may approach the excitation frequency (100 kHz). As a result, in narrow intervals of field around integer fillings we detect both Q_{in} and Q_{out} components; that is, the signal has a non-zero phase shift. But at fields somewhat higher or lower than this interval, no detailed contrast is observed in Q_{out} but much is seen in Q_{in} . In Fig. 4 we display this behaviour by showing Q_{in} and Q_{out} images for two different magnetic fields, with the higher field (3.005 T) closer to integer Landau level filling factor.

A more elaborate model treats the 2DES as a distributed RC system with no unique roll-off frequency. Consider a region of size L , ranging from the size of the sample L_{max} (several millimetres) to the resolution limit given by L_{min} (900 Å). Given a sheet of uniform resistivity in two dimensions, the effective resistance of this region does not depend on its size, but the self capacitance increases linearly with L . The roll-off frequency depends on the size L of the region in question: $f_0(L) = 1/(2\pi RL)$. Thus, $f_0(L)$ increases as L diminishes. Therefore, when the roll-off frequency becomes of the order of the measurement frequency, the charging of smaller length scales still behaves as in the low-frequency limit, and features charge in-phase with their surroundings. Thus only compressibility features can be detected at small length scales.

Why is the contrast due to compressibility variation seen in such a narrow interval of magnetic fields? The simplest answer is that the features may exist only near integer fillings. Another possibility is that the contrast is enhanced near integer fillings because of the lack of screening by the 2DES. This enhancement will occur when the screening radius in the 2DES becomes of the order of the distance to the tip, that is, in a narrow interval around an integer filling.

We now consider what properties of the quantum Hall effect could give rise to the observed structures. From measurements of the sample capacitance versus magnetic field, we know that $\nu = 4$ occurs at 3.040 T in the bulk. The droplets (see schematic in Fig. 3) seen at 3.000 T and 3.005 T can be explained as regions of higher electron density than their surroundings, which contain electrons in Landau levels $\nu > 4$. As the magnetic field strength is increased, the higher Landau levels depopulate, and the droplets shrink in size²⁶ and eventually disappear altogether.

At lower fields, the droplets are larger and at 2.995 T appear to overlap with a relatively bright 'island'. We thus surmise that the island is also a region of higher electron density where $\nu > 4$. Filaments border this area and thus occur at positions where there is a density gradient. We note that we have never observed filaments which terminate: either they form closed loops or meander all the way across the images.

To understand the origin of filaments one has to explain clumping of the electron density in a 2DES where the main interaction between electrons is Coulomb repulsion. It is known that in some cases the exchange interaction between electrons can overcome the direct density–density interaction and change the effective interaction from repulsion to attraction. There are several models that use exchange and predict filamentary structures in the quantum Hall system. In particular, Chamon and Wen²⁷ and MacDonald²⁸ have described charge structures forming near the edges of a 2DES in the presence of a large density gradient. Recently, Koulakov, Fogler and Shklovskii²⁹ considered a 2DES containing slightly greater than integer number of filled Landau levels, which is the same situation as Fig. 3. Using the Hartree–Fock method, Koulakov *et al.* found that the electrons in the higher level tended to form charge density waves; that is, a family of equally spaced filaments.

When the "island" (Fig. 3) is observed as a function of field, it is clear that unlike the "droplets", this region does not shrink but rather expands with increasing magnetic field. This at first seems inconsistent with the idea that the island contains electrons in a higher Landau level which depopulates as the field is raised.

However, as the field increases, more dark filaments appear towards the centre of the island. In general, as the field is increased, the filaments widen into dark and bright bands moving outwards from the centre of the island. The images suggest that some regions of higher ν depopulate in a different fashion to that considered previously. In the island, the density of electrons in the higher Landau level is reduced by 'stretching' the area. The electron density in the island can be thought of as an accordion, and the spacing between bands of high electron density increases with increasing magnetic field.

We have shown that the SCA technique has a profound capability to look into systems that have been inaccessible to imaging with comparable resolution. Even though it is not at present possible to determine the exact origin of the structures, there is no question that they exist. The observed shapes, rapidly evolving in magnetic field, directly mirror the electronic structures in the highly correlated Hall fluid. The range of experimental conditions which produce such shapes, the importance of disorder, and the nature of the fractional Hall regime will be objects of further study. $\square$

Received 4 September; accepted 22 December 1997.

1. Prange, R. E. & Girvin, S. M. (eds) *The Quantum Hall Effect* 2nd edn (Springer, New York, 1990).
2. Das Sarma, S. & Pinczuk, A. (eds) *Perspectives in Quantum Hall Effects* (Wiley Inter-Science, New York, 1997).
3. Laughlin, R. B. Quantized Hall conductivity in two dimensions. *Phys. Rev. B* **23**, 5632–5634 (1981).
4. Halperin, B. I. Quantized Hall conductance, current-carrying edge states, and the existence of extended states in a two-dimensional disordered potential. *Phys. Rev. B* **25**, 2185–2188 (1982).
5. Iordansky, S. V. On the conductivity of two dimensional electrons in a strong magnetic field. *Solid State Commun.* **43**, 1–3 (1982).
6. Chalker, J. T. The quantum Hall effect: a sum rule. *Surf. Sci.* **142**, 182–185 (1984).
7. von Klitzing, K., Dorda, G. & Pepper, M. New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Phys. Rev. Lett.* **45**, 494–497 (1980).
8. Tsui, D. C., Stormer, H. L. & Gossard, A. C. Zero-resistance state of two-dimensional electrons in a quantizing magnetic field. *Phys. Rev. B* **25**, 1405–1407 (1982).
9. Goldberg, B. B. *et al.* Optical investigations of the integer and fractional quantum Hall effects: energy plateaus, intensity minima, and line splitting in band-gap emission. *Phys. Rev. Lett.* **65**, 641–644 (1990).
10. Kukushkin, I. V. *et al.* Evidence of the triangular lattice of crystallized electrons from time resolved luminescence. *Phys. Rev. Lett.* **72**, 3594–3597 (1994).
11. Eisenstein, J. P. *et al.* Density of states and de Haas-van Alphen effect in two-dimensional electron systems. *Phys. Rev. Lett.* **55**, 875–878 (1985).
12. Ashoori, R. C. & Sillsbee, R. H. The Landau level density of states as a function of Fermi energy in the two-dimensional electron gas. *Solid State Commun.* **81**, 821–825 (1992).
13. Smith, T. P., Goldberg, B. B., Stiles, P. J. & Heiblum, M. Direct measurement of the density of states of a two-dimensional electron gas. *Phys. Rev. B* **32**, 2696–2699 (1985).
14. Eriksson, M. A., Beck, R. G., Topinka, M. & Katine, J. A. Cryogenic scanning probe characterization of semiconductor nanostructures. *Appl. Phys. Lett.* **69**, 671–673 (1996).
15. Yoo, M. J. *et al.* Scanning single-electron transistor microscopy: imaging individual charges. *Science* **276**, 579–582 (1997).
16. Williams, C. C., Slinkman, J. & Hough, W. P. Lateral dopant profiling with 200 nm resolution by scanning capacitance microscopy. *Appl. Phys. Lett.* **55**, 1662–1664 (1989).
17. Kochanski, G. P. Nonlinear alternating-current tunneling microscopy. *Phys. Rev. Lett.* **62**, 2285–2288 (1989).
18. Stranick, S. J. & Weiss, P. S. A tunable microwave frequency alternating current scanning tunneling microscope. *Rev. Sci. Instrum.* **65**, 918–921 (1994).
19. Bourgoin, J.-P., Johnson, M. B. & Michel, B. Semiconductor characterization with the scanning surface harmonic microscope. *Appl. Phys. Lett.* **65**, 2045–2047 (1994).
20. Shockley, W., Queisser, H. J. & Hooper, W. W. Charges on oxidized silicon surfaces. *Phys. Rev. Lett.* **11**, 489–490 (1963).
21. Efros, A. L. Non-linear screening and the background density of states of 2DEG in magnetic field. *Solid State Commun.* **67**, 1019–1022 (1988).
22. Chklovskii, D. B. & Lee, P. A. Transport properties between quantum Hall plateaus. *Phys. Rev. B* **48**, 18060–18078 (1993).
23. Trugman, S. A. Localization, percolation, and the quantum Hall effect. *Phys. Rev. B* **27**, 7539–7546 (1983).
24. Pruisken, A. M. M. Universal singularities in the integral quantum Hall effect. *Phys. Rev. Lett.* **61**, 1297–1300 (1988).
25. Hukenshtein, B. Scaling theory of the integer quantum Hall effect. *Rev. Mod. Phys.* **67**, 357–396 (1994).
26. McEuen, P. L. *et al.* Self-consistent addition spectrum of a Coulomb island in the quantum Hall regime. *Phys. Rev. B* **45**, 11419–11422 (1992).
27. Chamon, C. d. C. & Wen, X. G. Sharp and smooth boundaries of quantum Hall liquids. *Phys. Rev. B* **49**, 8277–8281 (1994).
28. MacDonald, A. H., Yang, S. R. E. & Johnson, M. D. Quantum dots in strong magnetic fields: stability criteria for the maximum density droplet. *Aust. J. Phys.* **46**, 345–358 (1993).
29. Koulov, A. A., Fogler, M. M. & Shklovskii, B. I. Charge density wave in two-dimensional electron liquid in magnetic field. *Phys. Rev. Lett.* **76**, 499–502 (1996).
30. Ashoori, R. C. *et al.* Single-electron capacitance spectroscopy of discrete quantum levels. *Phys. Rev. Lett.* **68**, 3088–3091 (1992).

Acknowledgements. We thank A. Cohen, E. Atmaca, M. Brodsky, H. B. Chan, A. Folch, S. Heemeyer, A. Shytov, D. Silevitch and N. B. Zhitenev for technical help, and B. I. Halperin, H. F. Hess, R. B. Laughlin, P. A. Lee and B. I. Shklovskii for discussions. This work was supported by the Office of Naval Research, the Packard Foundation, JSEP, and the National Science Foundation DMR.

Correspondence and requests for materials should be addressed to R.C.A. (e-mail: ashoori@MIT.EDU).

Highly mobile electrons and holes on isolated chains of the semiconducting polymer poly(phenylene vinylene)

Romano J. O. M. Hoofman, Matthijs P. de Haas, Laurens D. A. Siebbeles & John M. Warman

IRI, Delft University of Technology, Mekelweg 15, 2629 JB Delft, The Netherlands

The nature of the charge carriers in 'conducting' polymers is of considerable interest at present^{1,2}, largely on the basis of the technological potential of these materials for use as the semiconducting layer in field-effect transistors (FETs) and the emissive layer in light-emitting diodes³ (LEDs). One of the main outstanding questions concerns the relative importance of intra-versus inter-chain charge transfer in determining the overall rate of charge transport. Here we apply the pulse-radiolysis time-resolved microwave conductivity technique⁴ to dilute solutions of a soluble dialkoxy derivative of the semiconducting polymer poly(phenylene vinylene), PPV, by which means we determine the one-dimensional intra-chain mobilities of electrons and holes on isolated polymer chains free from inter-chain interactions. The values so obtained—0.5 and 0.2 cm² V⁻¹ s⁻¹ respectively—are considerably larger than the mobilities measured previously for bulk PPV-based materials^{5–9}. This suggests that considerable improvement in the performance characteristics (in particular switching time and maximum current) of organic FET and LED devices should be possible if material purity and structural order can be better controlled.

In this work, we used a dilute solution in benzene of poly(2-methoxy-5-[2'-ethyl-hexyloxy]-phenylene vinylene), 'MEH-PPV'. The phenylene vinylene (PV) repeat-unit concentration, as determined by optical absorption, was 2×10^{-3} M. This corresponds to a concentration of 2×10^{-6} M of polymer chains with an average length of $\sim 1,000$ PV units based on a relative molecular mass of $200,000 \pm 100,000$. We have found no evidence of aggregation occurring in the benzene solution of MEH-PPV used. Certain other dialkoxy-PPVs have been found to have varying tendencies to aggregate in solution, as evidenced by gel formation, the development of new adsorption and emission bands^{10,11}, and in particular a very marked increase in the photoconductivity due to the occurrence of inter-chain charge transfer in aggregated systems¹¹. We observed none of these effects for the present MEH-PPV solution, even after it had been standing in a refrigerator for several weeks.

Charge carriers are created uniformly throughout the solution by pulsed ionization using 5-ns pulses of 3-MeV electrons from a Van de Graaff accelerator; these electrons have a penetration depth (15 mm) much larger than the 3.5 mm thickness of the sample. Energy is initially absorbed from the beam mainly by the solvent, resulting in the formation of benzene radical cations ('holes'), Bz⁺, and excess electrons, e⁻. We calculated the concentration of electron-hole pairs formed in the pulse (using the known free ion yield of 0.053 per 100 eV absorbed^{12,13} and the accurately measured dose) to be $\sim 1 \times 10^{-7}$ M, that is, considerably less than the concentration of polymer chains in the solution.

Liquid benzene has an electron affinity¹⁴ of ~ 0.1 eV, which is much lower than the ~ 2.5 eV estimated¹⁵ for PPV in the condensed phase. The excess electrons initially formed in the solvent would therefore be expected to react with the polymer chains in the solution according to:



Similarly, the fact that the ionization potential of benzene (9.2 eV; ref. 16) is much higher than the ~ 7 eV for PPV (ref. 15; gas-phase values) suggests that electron abstraction by Bz^+ from PPV, reaction (2), should also readily occur.



So in this way, the primary solvent charge carriers will be converted into electrons and holes on the polymer chains. Because a much smaller concentration of e^- and Bz^+ pairs are formed in the pulse than there are PPV chains in the solution, single-electron attachment or abstraction events per chain will predominate; this will result, after reaction, in a solution containing mainly PPV^- and PPV^+ .

To investigate the mobility of charge carriers within the solution, we use the time-resolved microwave conductivity technique. This electrodeless method involves propagation of microwaves through the sample and monitoring, with nanosecond time-resolution, of changes in the transmitted power which occur as a result of pulsed ionization. The parameter derived from the experimental results is the transient change in conductivity of the sample, $\Delta\sigma(t)$, which is given by

$$\Delta\sigma(t) = e\sum N_i(t)\mu_i \quad (3)$$

where e is the elementary charge, $N_i(t)$ is the time-dependent concentration of a given charged species i and μ_i is its mobility.

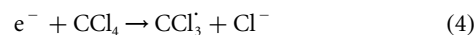
The microwave frequency-band used was 26–38 GHz, corresponding to a reciprocal radian frequency, $1/2\pi f$, of 5 ps. The average microwave power level was 100 mW which corresponds to an electric-field amplitude within the sample of ~ 25 V cm $^{-1}$. The solution was contained in a cell consisting of a gold-plated 27-mm-long section of waveguide (cross-section 3.5×7.1 mm) that was short-circuited at one end by a metal plate and closed at the other end by a polyimide film. The response time of detection was ~ 1 ns. Deaeration and addition of second solutes was carried out on a vacuum line. Full details of the microwave circuitry and data reduction procedures involved in the time-resolved microwave conductivity technique have been published elsewhere⁴.

The conductivity transient that we observed for pure benzene is shown in Fig. 1, top. From the end-of-pulse conductivity, and the

known concentration of charge carriers, we calculate the sum of the e^- and Bz^+ mobilities to be 0.15 cm 2 V $^{-1}$ s $^{-1}$. This is very close to the value of 0.13 cm 2 V $^{-1}$ s $^{-1}$ determined by a d.c. time-of-flight method for the mobility of the excess electron^{13,14} in benzene. The mobility of Bz^+ , which is thought to be in equilibrium with the dimer ion, Bz_2^+ (ref. 17), is much lower at $\sim 1 \times 10^{-3}$ cm 2 V $^{-1}$ s $^{-1}$ and is controlled mainly by molecular diffusion. The decay of the conductivity in 'pure' benzene on a timescale of ~ 100 ns is attributed to the attachment of excess electrons to impurities still present in the solvent.

Also shown in Fig. 1 top is the dramatically different conductivity transient that we found for the MEH-PPV solution. Here, a growth in conductivity is actually observed after the pulse, and the decay occurs on a timescale almost three orders of magnitude longer than for the pure solvent. The after-pulse growth to a level even larger than in benzene alone is an immediate qualitative indication that the sum of the mobilities of the electron and hole after transfer to the polymer chains via reactions (1) and (2) is in fact higher than that of the initial solvent charge carriers; that is, $[\mu(e_{PPV}^-) + \mu(h_{PPV}^+)] > 0.15$ cm 2 V $^{-1}$ s $^{-1}$. The symbols $\mu(e_{PPV}^-)$ and $\mu(h_{PPV}^+)$ are used to distinguish the mobility of a charge carried on the polymer backbone from the mobility of the charged polymer chain as a whole, which will be denoted $\mu(PPV^-)$ and $\mu(PPV^+)$. The latter will be controlled by molecular diffusion and will be consequently orders of magnitude lower than the value given above.

To discriminate between the electron and hole contributions to the conductivity in the polymer solution, we added excess concentrations of the competitive electron and hole scavengers CCl_4 and NH_3 which react according to:



The influence of these scavengers on the conductivity transient for benzene alone, that is, without MEH-PPV, is shown in Fig. 1 top. In the solution containing CCl_4 , the conductivity is reduced dramatically owing to conversion of the highly mobile excess electrons to much less mobile chloride ions. On the other hand, the conductivity is little influenced by the presence of NH_3 because this solute is unreactive towards excess electrons but does react with the low-mobility Bz^+ ions.

When excess concentrations ($\sim 10^{-2}$ M) of CCl_4 or NH_3 are added to the polymer containing solution, we found a substantial reduction in the conductivity signal, as shown by a comparison of the polymer solution in the upper section of Fig. 1 with the transients in the lower section of this figure. But in both cases, large conductivity transients remain, showing that both the electron and the hole are highly mobile on a polymer chain. The much slower growth of the signal in the CCl_4 /PPV solution results from the low mobility of Bz^+ and the corresponding low value of the rate constant for the abstraction reaction, (2), which leads to the formation of PPV^+ . The much more highly mobile electrons in the NH_3 solution, on the other hand, diffuse to and react with the polymer chains much more rapidly, resulting in a faster growth of this conductivity component. The fact that the conductivity is initially close to constant in the NH_3 /PPV solution is a qualitative indication that the mobility of electrons on the polymer chains (that is, $\mu(e_{PPV}^-)$) must be very similar to that of the excess electrons they replace, 0.15 cm 2 V $^{-1}$ s $^{-1}$.

We found that the decay rate of the conductivity transient in the CCl_4 /PPV solution increases with increasing dose in the pulse, that is, increasing charge-carrier concentration. This indicates that the decay of PPV^+ involves the neutralization of PPV^+ with the counterion, Cl^- . The more rapid decay of PPV^- in the NH_3 /PPV solution is independent of dose, and probably results from reaction of electrons with impurity trapping centres intrinsically present as more electronegative defects on the polymer chain or impurities in the solution.

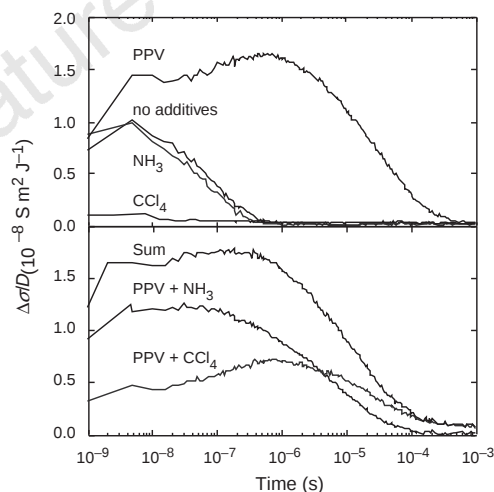


Figure 1 Microwave conductivity transients observed on 5-ns pulsed ionization of benzene solutions. Top panel: no additives, 1×10^{-2} M of the electron scavenger CCl_4 , 3×10^{-3} M of the positive-ion scavenger NH_3 , and $\sim 2 \times 10^{-6}$ M PPV polymer chains. Bottom panel: $\sim 2 \times 10^{-6}$ M PPV together with either 1×10^{-2} M CCl_4 or 3×10^{-3} M NH_3 . The sum of these transients is also shown, for comparison with the PPV solution without added scavenger shown in the top panel. The vertical axis is the radiation-induced conductivity, $\Delta\sigma$, normalized by the absorbed dose, D , in the pulse.

From quantitative kinetic fits to the absolute magnitude of the conductivity transients observed in the PPV/NH₃ and PPV/CCl₄ solutions, and using the known concentration of charge carriers formed, we calculated values of $\mu(e_{\text{PPV}}^-) = 0.15 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and $\mu(h_{\text{PPV}}^+) = 0.06 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. As the polymer chains are orientated randomly within the solution, these isotropic mobility values correspond to one-dimensional, intra-chain mobilities a factor of 3 higher, or 0.5 and $0.2 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, respectively. These values are similar to those found for single crystals of aromatic compounds¹⁸.

If charge transport along the polymer chains is free from static energetic disorder, then the high-frequency mobility determined in the present work should be equal to the low-field d.c. mobility, that is, the mobility that would be determined in a hypothetical experiment using nano-electrodes attached to the polymer chain ends. The recent experiments of Tans *et al.*¹⁹ on single carbon nanotubes show that such experiments no longer belong to the realms of fantasy.

In terms of a one-dimensional hopping model of charge transport, μ is related to the average time between random displacements or 'jumps' of charged centres, τ , and the average displacement per jump, λ , by

$$\mu = \frac{\lambda^2 e}{2k_B T \tau} \quad (6)$$

where k_B is the Boltzmann constant and T is the temperature. Taking the average displacement to be equal to the length of a PPV monomer unit (that is, $\lambda \approx 7 \text{ \AA}$), the values of τ corresponding to the electron and hole mobilities determined are 0.2 and 0.5 ps, respectively.

If dynamic fluctuations in the conformation of the polymer chains occur on a timescale longer than the jump time, then quasi-static-disorder effects could become operative and cause the d.c. mobility to be lower than the high-frequency value. Using a Monte Carlo computational technique described previously²⁰ we have calculated the effect of energetic disorder on the intra-chain mobility for d.c. and 30-GHz field conditions. These calculations show that the d.c. mobility would be at most a factor of ~ 2 lower than the value measured at 30 GHz.

We conclude that mobilities larger than $0.1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and even approaching $1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ should in principle be possible in constant-voltage device structures based on poly(phenylene vinylene) derivatives. Values of this order of magnitude have in fact been derived for thin layers of PPV derivatives, using short-timescale photoconductivity measurements based on the assumption of a quantum yield of charge carriers of 0.01 (refs 21–23). However, direct time-of-flight measurements have yielded lower values close to $1 \times 10^{-4} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (refs 5–7) and even lower values have been estimated from the current/voltage characteristics of device structures^{8,9}. Such large differences cannot be ascribed simply to the more pronounced influence of static disorder in the bulk solid materials. The main factors are undoubtedly the barriers to charge transport presented by domain and grain boundaries and, in the case of the electron, the presence of deep trapping sites²⁴.

The present results indicate that by paying more attention to structural integrity and chemical purity, considerable improvements in the performance of organic molecular devices should be possible. This would include faster switching times and higher operating currents; such improvements are essential for the commercial applications envisaged. $\square$

Received 18 July; accepted 29 December 1997.

- Conwell, E. M. in *Handbook of Organic Conductive Molecules and Polymers* (ed. Nalwa, H. S.) 1–45 (Wiley, New York, 1997).
- Epstein, A. J. Electrically conducting polymers: science and technology. *Mater. Res. Soc. Bull.* **22**(6), 16–24 (1997).
- Yang, Y. Polymer electroluminescent devices. *Mater. Res. Soc. Bull.* **22**(6), 31–38 (1997).
- Infelta, P. P., de Haas, M. P. & Warman, J. M. The study of the transient conductivity of pulse irradiated dielectric liquids on a nanosecond timescale using microwaves. *Radiat. Phys. Chem.* **10**, 353–365 (1977).
- Meyer, H., Haarer, D., Naarmann, H. & Hörhold, H. H. Trap distribution for charge carriers in poly(paraphenylene vinylene) (PPV) and its substituted derivative DPOP-PPV. *Phys. Rev. B* **52**, 2587–2598 (1995).

- Kryukov, A. Y., Saidov, A. C. & Vannikov, A. C. Charge carrier transport in poly(phenylene vinylene) films. *Thin Solid Films* **209**, 84–91 (1992).
- Gailberger, M. & Bässler, H. DC and transient photoconductivity of poly(2-phenyl-1,4-phenylenevinylene). *Phys. Rev. B* **44**, 8643–8651 (1991).
- Blom, P. W. M., de Jong, M. J. M. & Vlegaar, J. J. M. Electron and hole transport in poly(p-phenylene vinylene) devices. *Appl. Phys. Lett.* **68**, 3308–3310 (1996).
- Pichler, K., Jarrett, C. P., Friend, R. H., Ratier, B. & Moliton, A. Field-effect transistors based on poly(p-phenylene vinylene) doped by ion implantation. *J. Appl. Phys.* **77**, 3523–3527 (1995).
- Peeters, E., Delmotte, A., Janssen, R. A. J. & Meijer, E. W. Chiroptical properties of poly[2,5-bis(S)-methylbutoxy]-1,4-phenylene vinylene. *Adv. Mater.* **9**, 493–496 (1997).
- Gelinck, G. H., Warman, J. M. & Staring, E. G. J. Polaron pair formation, migration and decay on photoexcited poly(phenylenevinylene) chains. *J. Phys. Chem.* **100**, 5485–5491 (1996).
- Schmidt, W. F. & Allen, A. O. Free-ion yields in sundry irradiated liquids. *J. Chem. Phys.* **52**, 2345–2350 (1970).
- Gee, N. & Freeman, G. R. Electron thermalization distances and free-ion yields in dense gaseous and liquid benzene. *Can. J. Chem.* **70**, 1618–1622 (1992).
- Itoh, K. & Holroyd, R. Effect of pressure on the electron mobility in liquid benzene and toluene. *J. Phys. Chem.* **94**, 8850–8854 (1990).
- Meyers, F., Heeger, A. J. & Brédas, J.-L. Fine tuning of the band gap in conjugated polymers via control of block copolymer sequences. *J. Chem. Phys.* **97**, 2750–2758 (1992).
- Ionisation Potentials, Appearance Potentials and Heats of Formation of gaseous Positive Ions* (NSRDS-NBS Rep. No. 26, US Govt Printing Office, Washington DC, 1969).
- Inokuchi, Y. *et al.* Formation of benzene dimer cations in neat liquid benzene studied by femtosecond transient absorption spectroscopy. *Chem. Phys. Lett.* **269**, 298–304 (1997).
- Karl, N. in *Landolt-Börnstein, New Series, Group III Vol. 17, Semiconductors Subvol. i* (Springer, Berlin, 1985).
- Tans, S. J. *et al.* Individual single-wall nanotubes as quantum wires. *Nature* **386**, 474–477 (1996).
- Hilt, O. & Siebbeles, L. D. A. Time and frequency dependent charge carrier mobility of one-dimensional chains with energetic disorder. *Chem. Phys. Lett.* **269**, 257–262 (1997).
- Shen, Y. Q., Lindenberger, H., Bleier, H. & Roth, S. in *Electronic Properties of Conjugated Polymers III* (eds Kuzmany, H., Mehring, M. & Roth, S.) 96–99 (Springer, Berlin, 1989).
- Bradley, D. D. C., Shen, Y. Q., Bleier, H. & Ruth, S. Transient photoconductivity in highly oriented poly(p-phenylenevinylene). *J. Phys. C* **21**, L515–L522 (1988).
- Lee, C. H., Yu, G., Moses, D. & Heeger, A. J. Picosecond transient photoconductivity in poly(p-phenylenevinylene). *Phys. Rev. B* **49**, 2396–2407 (1994).
- Antoniadis, H., Abkowitz, M. A. & Hsieh, B. R. Carrier deep-trapping mobility-lifetime products in poly(p-phenylenevinylene). *Appl. Phys. Lett.* **65**, 2030–2032 (1994).

Acknowledgements. We thank D. H. Hwang and A. B. Holmes (Melville Laboratory for Polymer Synthesis, Cambridge, UK) for supplying MEH-PPV.

Correspondence and requests for materials should be addressed to M.P.d.H. (e-mail: mhaas@iri.tudelft.nl).

Uniform quantum-dot arrays formed by natural self-faceting on patterned substrates

Richard Nötzel, Zhichuan Niu, Manfred Ramsteiner, Hans-Peter Schönherr, Achim Tranpert, Lutz Däweritz & Klaus H. Ploog

Paul-Drude-Institut für Festkörperelektronik, Hausvogteiplatz 5-7, D-10117 Berlin, Germany

Of the approaches currently under investigation for the fabrication of functional III–V semiconductor nanostructures, self-organized growth mechanisms and directed growth on patterned substrates have yielded quantum wires and dots with the best structural and electronic properties¹. In patterned growth, high densities of structures are difficult to obtain; self-organization, on the other hand, can provide densely packed structures with good crystal quality, but generally offers limited control over nanostructure uniformity and spatial position. In the case of quantum dots, non-uniformity of size and shape is clearly undesirable, as the resulting structures will exhibit a broad range of electronic and optical properties, effectively smearing out the sought-for zero-dimensional behaviour of the dot ensemble. Here we demonstrate a method for improving size uniformity, while maintaining a high density of quantum dots, that combines elements of both self-organization and patterning. The photoluminescence spectrum of the resulting ordered arrays of quantum dots is dominated by a single sharp line, rather than the series of sharp lines that would indicate transitions in quantum dots of different sizes^{2–7}.

Starting from a two-dimensional GaAs/(AlGa)As quantum well grown by molecular beam epitaxy (MBE) on patterned GaAs (311)A (Ga-terminated) substrates, we have recently found the evolution of a fast-growing sidewall for mesa stripes orientated along the $[01\bar{1}]$ direction⁸. Preferential migration of Ga atoms from the mesa top and mesa bottom towards the sidewall increases the thickness of the GaAs layer used to form the quantum well at the foot of each mesa stripe, thus reducing the dimensionality to one. For mesa heights in the quantum size regime (10–15 nm), this new growth mode allows the fabrication of quasi-planar lateral quantum wires, several tens of

nanometres wide, which are bound by a very smooth convex curved surface profile without faceting. The high structural perfection of the wires is shown in excellent optical properties with narrow photoluminescence (PL) linewidths, high PL efficiency and clearly resolved one-dimensional sub-bands in photoluminescence excitation (PLE) spectroscopy⁹. On the other hand, if atomic hydrogen is added during MBE on GaAs (311)A substrates, natural one-dimensional surface corrugations on the nanometre scale form in the $[\bar{2}33]$ direction—that is, perpendicular to the strip direction mentioned above. Our approach is to combine these two phenomena to pattern the GaAs quantum-well layer in two dimensions, to give rise to quantum dots as shown schematically in Fig. 1.

The influence of atomic hydrogen on the surface morphology of GaAs grown on planar GaAs (311)A substrates is shown in the atomic force microscopy (AFM) images in Fig. 2a, b. In conventional MBE, the surface morphology is smooth on the mesoscopic scale with micrometre-sized islands. Some areas on the surface show a shallow, irregular fine structure of steps extending several tens of nanometres. In the presence of atomic hydrogen, distinct quasi-periodic arrays of symmetrically arranged upward and downward steps are formed along $[\bar{2}33]$. The lateral periodicity of the step array is ~ 40 nm and the step height is 2–3 nm. The step array is straight over several micrometres with very few crossing points of the step edges, and can thus be used as a template for the self-organizing formation of wire-like nanostructures. The surface morphology observed in atomic-hydrogen-assisted MBE is similar to that found during metal organic vapour phase epitaxy (MOVPE)¹⁰. The lateral periodicity and step heights are comparable for the same growth conditions, thus revealing a similar growth mechanism. Atomic hydrogen, which is present in MOVPE owing to decomposition of

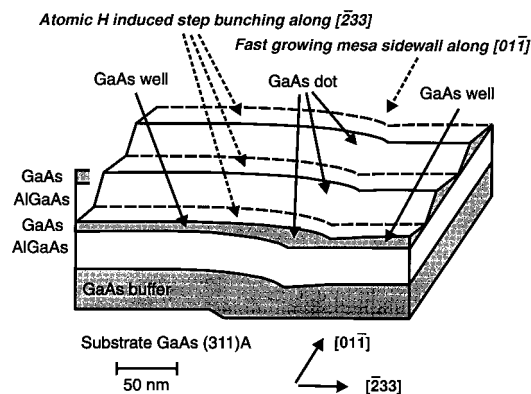


Figure 1 Schematic illustration of the quantum-dot array. This was formed by atomic-hydrogen-assisted MBE on patterned GaAs (311)A substrates at the sidewall of mesa stripes along $[01\bar{1}]$.

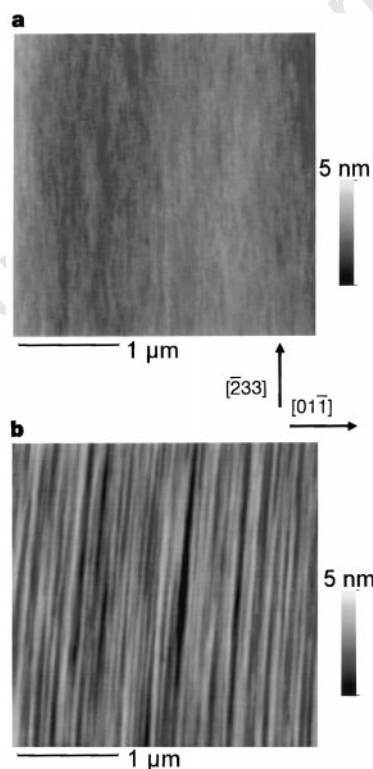


Figure 2 AFM images of the surface of 100-nm GaAs grown on a GaAs (311)A substrate. **a**, Grown by conventional MBE; **b**, grown by atomic-hydrogen-assisted MBE showing the natural periodic step array along $[\bar{2}33]$. The growth rate was $0.3 \mu\text{m h}^{-1}$ at a temperature of 620°C . Atomic hydrogen was supplied by a cracker cell using a tungsten filament heated to $\sim 1,800^\circ\text{C}$.

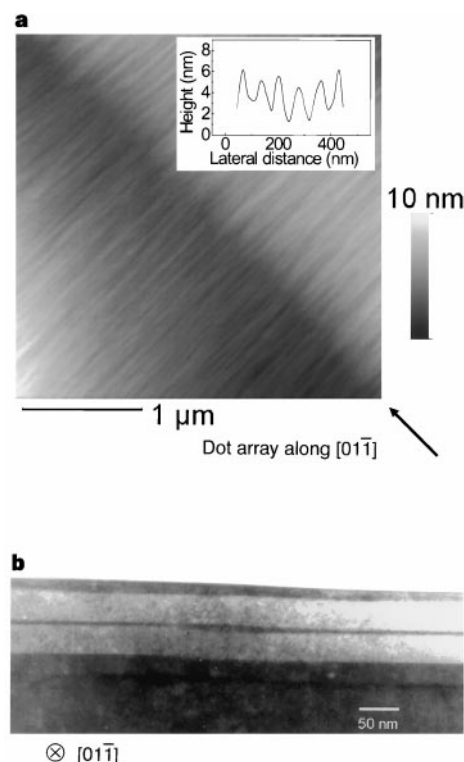


Figure 3 Structural analysis of the quantum-dot array. **a**, AFM image of the surface morphology of the convex curved profile along the fast growing sidewall of mesa stripes along $[01\bar{1}]$ on patterned GaAs (311)A substrates overgrown by atomic-hydrogen-assisted MBE showing the natural step array along $[\bar{2}33]$ with a lateral periodicity of ~ 40 nm. The inset shows the AFM height profile along the bottom step corner. **b**, Cross-sectional TEM image viewed along $[01\bar{1}]$ of the 40–50-nm-wide quantum wire formed along the fast-growing sidewall in conventional MBE.

the precursors, is therefore thought to strongly promote step bunching of the microscopic surface corrugation in conventional MBE¹¹. The process, which is opposite to the flattening of (100) surfaces^{12,13} most probably relies on the passivation of free bonds¹⁴ which might increase the incorporation rate of Ga atoms at step-down sites¹⁵. But the atomic-scale mechanisms, including the microscopic arrangement of the surface corrugation and its reconstruction, are at present not well understood.

Using atomic-hydrogen-assisted MBE on patterned GaAs (311)A substrates, we prepared stripes (10–20 nm high) orientated along [011] by optical lithography and wet chemical etching⁸. The growth sequence consisted of a GaAs/(AlGa)As double heterostructure on a 50-nm-thick GaAs buffer layer capped with 20-nm GaAs. Figure 3a shows the AFM image of the GaAs surface at the sidewall of the mesa stripe along [011] after overgrowth with the typical convex curvature of the fast-growing sidewall. The surface profile is similar to the profile obtained in conventional MBE, and encloses a quantum-wire structure with a height of about twice the thickness of the adjacent quantum well and a width of 40–50 nm (see the cross-sectional transmission electron microscopy (TEM) image in Fig. 3b). With the supply of atomic hydrogen, the natural one-dimensional surface structure along $\bar{2}33$ due to step bunching is also established on the convex curved part of the sidewall (Fig. 3a). Importantly, no displacement of the surface corrugation occurs in the wire region at the bottom step corner, as is revealed by the spatial correlation of the step bunches continuing over the sidewall from the mesa top to the mesa bottom. The AFM height profile (in Fig. 3a inset) measured along the bottom step corner reveals the high lateral periodicity of the step bunches and height fluctuations in the range of one monolayer (however, these fluctuations might be overestimated in AFM owing to irregularities introduced by the native oxide and by surface contamination, which is unavoidable during measurements in air). This indicates the presence of dot-like nanostructures (see Fig. 1) in the corrugated GaAs layer which is embedded between the two lower and upper (AlGa)As barrier layers.

We took micro-photoluminescence spectra (Fig. 4a) at different positions from the sidewall of the GaAs/(AlGa)As quantum-dot heterostructure on the patterned substrate. The 3-nm-thick GaAs well layer grown with atomic hydrogen is covered by 50-nm-thick

Al_{0.7}Ga_{0.3}As barrier layers. Like the sidewall quantum wires grown by conventional MBE, the spectra show emission at higher energy from the corrugated quantum well close to the sidewall (and in the flat parts of the mesa structure) and also emission shifted to lower energy from the thicker GaAs region along the sidewall. The energy separation is comparable to that obtained for heterostructures with similar structural parameters grown by conventional MBE⁹. This confirms that in atomic-hydrogen-assisted MBE, the GaAs layer thickness at the sidewall (neglecting the periodic surface corrugation) is about twice the thickness of the adjacent GaAs well.

The emission from three-dimensionally confined excitons in quantum dots manifests itself in the very pronounced splitting of the micro-photoluminescence spectra into sharp lines. The emission from the corrugated quantum well (QW in Fig. 4) shows a series of peaks uniformly superimposed over the average spectra; these peaks are strongly position dependent owing to the localization of excitons at interface fluctuations which are random along the step bunches. The spectra are very similar to those of narrow quantum wells, quantum wires, or strained (InGa)As island structures, with the energy distribution of the sharp peaks reflecting the statistical fluctuations of the ground-state energy owing to random fluctuations in shape and size of the localization sites^{2–7}. In contrast, the spectral range of the emission from the quantum-dot arrays (QD in Fig. 4), always centred at 1.619 eV, is much narrower compared to the linewidth of the average spectrum. Figure 4b shows the micro-photoluminescence spectrum of the dot array with enhanced spectral resolution. The measured linewidth of the spectral peak, 60–70 μ eV, determined by the resolution of our spectrometer, proves the emission from quantum dots⁴. The spectrum is dominated by one single sharp peak without background emission over an energy range of 20–30 meV, indicating the high size-uniformity of the dot array. This is supported by the few weak features appearing in the spectra in Fig. 4a between the emission from the corrugated quantum well and the dot array; these features are attributed to the emission from several dots with different, random sizes which are easily detected—if only a small fraction of the dots were of equal size, a distinct, continuous background emission would occur; this is not observed. In the dot array, the random interface fluctuations along the step bunches are removed due to wire formation in patterned growth; this is the underlying principle of our approach. In other words, the periodic step bunching along the sidewall, due to the influence of atomic hydrogen, results in pronounced and regular thickness modulations in the sidewall-induced quantum wires to form the dot array characterized by the narrowing of the micro-photoluminescence spectra and a well defined emission energy position. Moreover, the spectra of the quantum-dot array remain unchanged when the diameter of the optical probing area is increased from 2 to 16 μ m (the largest opening of the pinhole), corresponding to the probing of several tens to several hundred dots. This area is already large enough for further processing without degradation of the optical quality. □

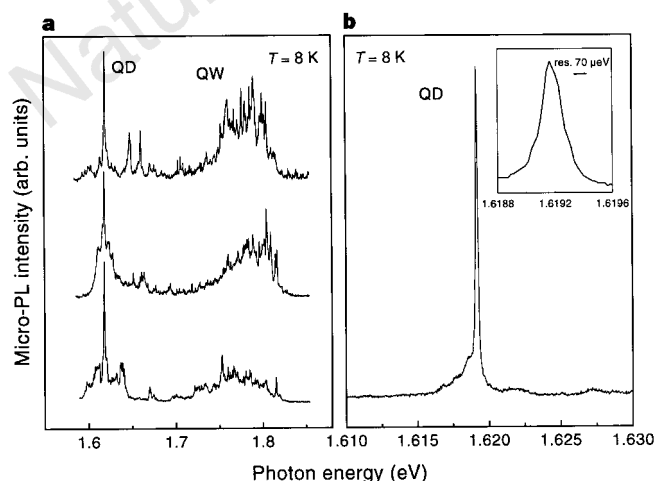


Figure 4 Optical properties of the quantum-dot array. **a**, Micro-photoluminescence (μ -PL) spectra taken at 8 K at different positions at the sidewall. QD and QW denote the emission from the quantum-dot array at the sidewall and from the adjacent corrugated quantum well, respectively. The diameter of the optical probing area is reduced by a confocal imaging system to 2 μ m. The excitation power of the Ar⁺-laser is 0.1 mW. **b**, High-resolution μ -PL spectra of the dot array at the sidewall.

Received 2 May; accepted 1 December 1997.

- Eberl, K., Petroff, P. M. & Demeester, P. (eds) *Low Dimensional Structures Prepared by Epitaxial Growth or Regrowth on Patterned Substrates* (NATO Advanced Sci. Inst. Ser. E., Vol. 298, Kluwer, Dordrecht, 1995).
- Hess, H. F., Betzig, E., Harris, T. D., Pfeiffer, L. N. & West, K. W. Near-field spectroscopy of the quantum constituents of a luminescent system. *Science* **264**, 1740–1745 (1994).
- Brunner, K., Abstreiter, G., Böhm, G., Tränkle, G. & Weimann, G. Sharp-line photoluminescence and two-photon absorption of zero-dimensional biexcitons in a GaAs/AlGaAs structure. *Phys. Rev. Lett.* **73**, 1138–1141 (1994).
- Gammon, D., Snow, E. S., Shanbrook, B. V., Kratzer, D. S. & Park, D. Homogeneous linewidths in the optical spectrum of a single gallium arsenide quantum dot. *Science* **273**, 87–90 (1996).
- Ramsteiner, M. *et al.* Influence of composition fluctuations in Al(Ga)As barriers on the exciton localization in thin GaAs quantum wells. *Phys. Rev. B* **55**, 5239–5242 (1997).
- Leon, R., Petroff, P. M., Leonard, D. & Fafard, S. Spatially resolved visible luminescence of self-assembled semiconductor quantum dots. *Science* **267**, 1966–1968 (1995).
- Marzin, J.-Y., Gerard, J.-M., Izrael, A., Barrier, D. & Bastard, G. Photoluminescence of single InAs quantum dots obtained by self-organized growth on GaAs. *Phys. Rev. Lett.* **73**, 716–719 (1994).
- Nötzel, R. *et al.* Selectivity of growth on patterned GaAs (311)A substrates. *Appl. Phys. Lett.* **68**, 1132–1134 (1996).

9. Richter, A. *et al.* Real-space transfer and trapping of carriers into single GaAs quantum wires studied by near-field optical spectroscopy. *Phys. Rev. Lett.* **79**, 2145–2148 (1997).
10. Nötzel, R., Temmyo, J. & Tammamura, T. Tunability of one-dimensional self-faceting on GaAs (311)A surfaces by metalorganic vapor phase epitaxy. *Appl. Phys. Lett.* **64**, 3557–3559 (1994).
11. Nötzel, R., Ledentsov, N., Däweritz, L., Hohenstein, M. & Ploog, K. Direct synthesis of corrugated superlattices on non-(100)-oriented surfaces. *Phys. Rev. Lett.* **67**, 3812–3815 (1991).
12. Okada, Y., Fujita, T. & Kawabe, M. Growth modes in atomic hydrogen assisted molecular beam epitaxy of GaAs. *Appl. Phys. Lett.* **67**, 676–678 (1995).
13. Okada, Y. & Harris, J. S. Jr Basic analysis of atomic-scale growth mechanisms for molecular beam epitaxy of GaAs using atomic hydrogen as a surfactant. *J. Vac. Sci. Technol. B* **14**, 1725–1728 (1996).
14. Miyamoto, Y. & Nonoyama, S. First principles calculation of molecular- and atomic hydrogen reactions on As-terminated GaAs (100) surfaces. *Phys. Rev. B* **46**, 6915–6921 (1992).
15. Gossmann, H. J., Sindén, F. W. & Feldman, L. C. Evolution of terrace size distributions during thin-film growth by step mediated epitaxy. *J. Appl. Phys.* **67**, 745–752 (1990).

Acknowledgements. We thank E. Wiebcke for preparing the patterned substrates. Part of this work was supported by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie.

Correspondence and requests for material should be addressed to R.N. (e-mail: notzel@pdi-berlin.de).

Atmospheric CO₂ concentration and millennial-scale climate change during the last glacial period

B. Stauffer*, T. Blunier*, A. Dällenbach*, A. Indermühle*, J. Schwander*, T. F. Stocker*, J. Tschumi*, J. Chappellaz†, D. Raynaud‡, C. U. Hammer‡ & H. B. Clausen‡

* Climate and Environmental Physics, Physics Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

† CNRS Laboratoire de Glaciologie (LGGE), BP 96, 38402 St Martin d'Hères Cedex, Grenoble, France

‡ Department of Geophysics, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen O, Denmark

The analysis of air bubbles trapped in polar ice has permitted the reconstruction of past atmospheric concentrations of CO₂ over various timescales, and revealed that large climate changes over tens of thousands of years are generally accompanied by changes in atmospheric CO₂ concentrations¹. But the extent to which such covariations occur for fast, millennial-scale climate shifts, such as the Dansgaard–Oeschger events recorded in Greenland ice cores during the last glacial period², is unresolved; CO₂ data from Greenland³ and Antarctic⁴ ice cores have been conflicting in this regard. More recent work suggests that Antarctic ice should provide a more reliable CO₂ record, as the higher dust⁵ content of Greenland ice can give rise to artefacts^{1,6,7}. To compare the rapid climate changes recorded in the Greenland ice with the global trends in atmospheric CO₂ concentrations as recorded in the Antarctic ice, an accurate common timescale is needed. Here we provide such a timescale for the last glacial period using the records of global atmospheric methane concentrations from both Greenland and Antarctic ice. We find that the atmospheric concentration of CO₂ generally varied little with Dansgaard–Oeschger events (<10 parts per million by volume, p.p.m.v.) but varied significantly with Heinrich iceberg-discharge events (~20 p.p.m.v.), especially those starting with a long-lasting Dansgaard–Oeschger event.

The $\delta^{18}\text{O}$ record of the GRIP ice core⁸ confirmed fast and drastic climate variations during the last glacial, so-called Dansgaard–Oeschger events^{2,8}. Over Greenland, rapid warmings of several degrees centigrade took place in a few decades, and were followed by a slower cooling and return to full glacial conditions. All but one of the Dansgaard–Oeschger events have a concomitant in ice-rafting debris recorded in sediment cores from the North Atlantic⁹. Simulations with ocean circulation models suggest that the fresh-

water input resulting from iceberg discharge has the potential to alter the North Atlantic Deep Water (NADW) formation^{10–12}. The effect of changes of the deep ocean circulation and of the marine biogeochemistry on the atmospheric CO₂ concentrations is an important open question.

Measurements of the CO₂ concentration in the air bubbles entrapped in ice cores from Greenland revealed large CO₂ variations, with levels of ~200 p.p.m.v. during the cold phases and ~250 p.p.m.v. during the mild phases of Dansgaard–Oeschger events³. Detailed measurements on the Byrd ice core (from Antarctica) did not show variations of this magnitude^{4,13}. It is now believed that the high CO₂ concentrations during mild phases in the Greenland ice cores do not represent the atmospheric concentration but are caused by an acid-carbonate reaction or by oxidation of organic material in the ice^{6,7}. Ice cores from Antarctica are less affected by such reactions^{1,6,7}.

Because the atmospheric CO₂ concentration can only be detected reliably in Antarctic ice, ice cores from both hemispheres are needed to investigate a possible correlation between fast climate changes in the Northern Hemisphere and the atmospheric CO₂ concentration. This requires precise synchronization of the age scales from Greenland and Antarctic ice cores. Sowers *et al.*¹⁴ showed that a synchronization of ice cores can be done using the global $\delta^{18}\text{O}$ signal of O₂. Another powerful tool is now available to synchronize the GRIP ice core from Greenland and the Byrd ice core from Antarctica; this tool is based on high-resolution methane records¹⁵. The accuracy of the synchronization is a factor of ten better with methane owing to the fast variations of the atmospheric CH₄ concentration. The variations are global, because the atmospheric residence time of CH₄ is

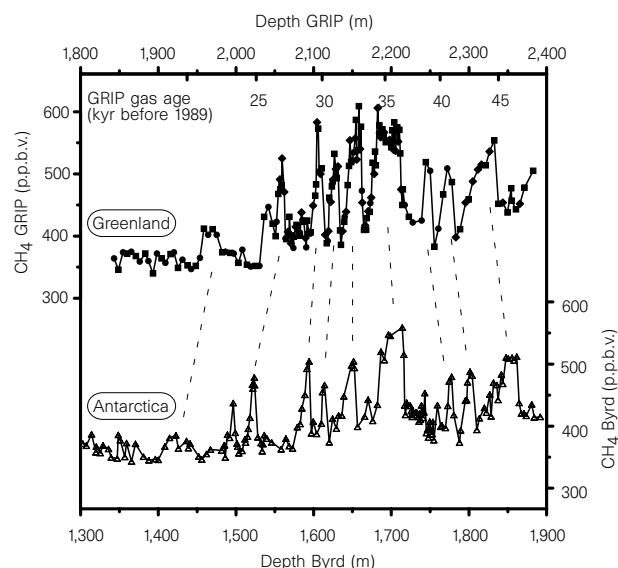


Figure 1 Methane concentrations measured on the GRIP and Byrd ice cores. Solid line with open triangles, Byrd station; solid line with filled symbols, GRIP (circles, published data¹⁷; squares, new data from the Physics Institute, University of Bern; diamonds, new data from the CNRS Laboratoire de Glaciologie, Grenoble). In both laboratories a melting–refreezing method was used to extract the gas from the ice^{16,17}. An ice sample was melted under vacuum in a glass container and then refrozen from the bottom. The extraction efficiency is close to 100%. The extracted air was then expanded into a sample loop and injected into a gas chromatograph equipped with a thermal conductivity detector (TCD) and a flame ionization detector (FID). The dashed lines denote peak values which are considered to characterize synchronous events. This identification is based on a Monte Carlo method maximizing the correlation between the two CH₄ records ($r^2 = 0.76$). The absolute timescale in kyr before 1989 (top) applies only to the GRIP record^{8,19}.

ten times larger than the interhemispheric exchange time¹⁶. Nevertheless, the interhemispheric CH₄ difference of ~5% may change slightly over time owing to changes in the latitudinal distribution of sources and sinks, as demonstrated for the Holocene period¹⁶. The methane records measured along the GRIP and the Byrd ice core show distinct variations, with a very similar pattern in the period 20–45 kyr before present, BP (Fig. 1). The covariance between CH₄ data from different sites confirms that the variations are of atmospheric origin. The GRIP methane variations parallel the Dansgaard–Oeschger events. During cold phases the concentration of methane is ~350 p.p.b.v. and during mild phases between 450 and 550 p.p.b.v. (ref. 17).

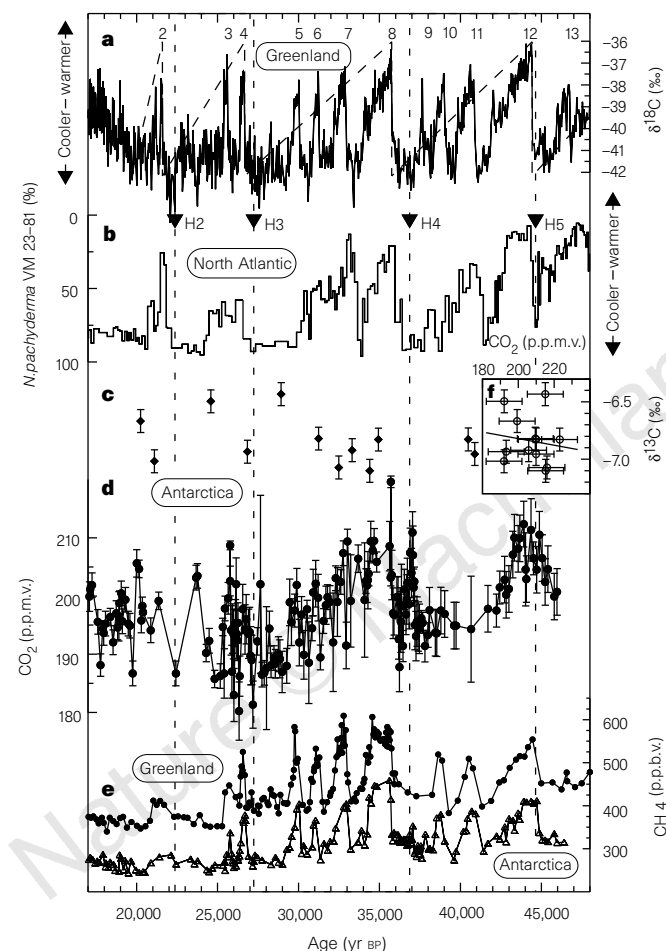


Figure 2 Comparison of CO₂ variations and climate changes. **a**, $\delta^{18}\text{O}$ values for the GRIP ice core⁸ together with Bond long-term cooling cycles, schematically shown as saw-tooth²⁰. **b**, *N. pachyderma* record from the deep sea core V23-81 (54° 15' N, 16° 50' W) as a proxy for sea surface temperature. Filled triangles show Heinrich events (ice-rifting events originating from the Hudson Strait)^{9,20}. **c**, $\delta^{13}\text{C}$ values measured on CO₂ extracted from Byrd ice samples²⁶. **d**, Byrd ice-core CO₂ values⁴ on the adjusted timescale. It is remarkable that the CO₂ concentration between Heinrich events 2 and 3, which shows only two narrow mild phases and a *N. pachyderma* abundance above 50%, remains smaller than between Heinrich events 3–4 and 4–5, where broad mild phases and *N. pachyderma* abundance below 50% are observed. **e**, Methane records of the GRIP and Byrd ice cores as in Fig. 1 but now plotted on the adjusted common timescale. Byrd values were lowered by 100 p.p.b.v. for better readability. **f**, $\delta^{13}\text{C}$ values plotted against CO₂ values measured on the same ice samples. The CO₂ measurements have been done volumetrically²⁶ which explains a higher uncertainty ($\sigma = \pm 10$ p.p.m.v.) compared to the results presented in **d**. The linear regression line shows an anticorrelation. A Monte Carlo simulation confirmed a very weak anticorrelation between $\delta^{13}\text{C}$ and CO₂ values. **a** and **b** are sections of Fig. 3 of Bond *et al.*^{9,20}.

Based on the fast CH₄ oscillations, a synchronization of the CH₄ signals is possible with an accuracy of about two centuries¹⁵ which is the typical timescale for methane increases at the beginning of a mild event. As timescale, we selected the one of the GRIP ice core^{8,18} for easier reference, but our conclusions do not depend on this choice. Because the occlusion of gases takes place typically at 70 m below the snow surface, the air in the bubbles is younger than the surrounding ice¹⁹. The age difference for GRIP varies between 350 and 1,290 yr with an uncertainty of ± 300 yr over the time range 47–17 kyr BP (ref. 19). The CH₄ records of both cores were matched using a Monte Carlo method, looking for maximum correlation between the two records¹⁹. Table 1 gives the depth as a function of the gas age for the GRIP and the Byrd ice core.

In Fig. 2, the Byrd CH₄ and CO₂ records with the adjusted Byrd timescale are shown, together with CH₄ and $\delta^{18}\text{O}$ records from the GRIP ice core. The age difference between ice and enclosed air is irrelevant for the comparison of only CH₄ and CO₂ records between two cores, and the accuracy of this synchronization is estimated to be ~200 yr. For the comparison of gas records with ice records, for example, the Byrd CO₂ record with the GRIP $\delta^{18}\text{O}$ record, the age difference between gases and ice is important. The uncertainty in this special case is only ~360 yr, because the age difference is well known for the GRIP ice core owing to a reliable timescale and precisely determined accumulation rates. A synchronization of the $\delta^{18}\text{O}$ records of the two cores could be achieved as soon as the age difference between gases and ice is determined for the Byrd core with a similar accuracy as for the GRIP core.

Together with the GRIP $\delta^{18}\text{O}$ record, the abundance of *Neogloboquadrina pachyderma* (s.) from the North Atlantic core V23-81 (refs 9, 20) (which is a proxy for sea surface temperature (SST)), and the position of the Heinrich layers are indicated in Fig. 2. The adjustment of the marine timescale has been done as in the original publication by Bond *et al.*²⁰.

The Byrd CO₂ concentration varies between about 180 and 210 p.p.m.v. during 47–17 kyr BP. We can therefore definitely exclude atmospheric CO₂ concentration variations parallel to Dansgaard–Oeschger events of the order of 50 p.p.m.v. as suggested by the measurements on Greenland ice cores³. The amplitudes of

Table 1 Depth versus age for ice-core samples

Gas age (yr BP)	GRIP depth (m)	Byrd depth (m)
17,000	1,854	1,284
18,000	1,889	1,317
19,000	1,916	1,350
20,000	1,937	1,383
21,000	1,958	1,410
22,000	1,979	1,421
23,000	1,994	1,432
24,000	2,010	1,446
25,000	2,026	1,472
26,000	2,046	1,504
27,000	2,068	1,530
28,000	2,080	1,552
29,000	2,093	1,573
30,000	2,112	1,594
31,000	2,125	1,613
32,000	2,140	1,633
33,000	2,162	1,653
34,000	2,173	1,673
35,000	2,192	1,702
36,000	2,216	1,722
37,000	2,225	1,742
38,000	2,237	1,763
39,000	2,253	1,780
40,000	2,263	1,793
41,000	2,281	1,805
42,000	2,290	1,817
43,000	2,305	1,832
44,000	2,323	1,850
45,000	2,338	1,867
46,000	2,347	1,885

Depths are given in m below surface. The uncertainty of the gas age relative to the age scale of the ice¹⁸ is ~300 yr for the GRIP core and ~360 yr for the Byrd core.

methane variations, tracking all Dansgaard–Oeschger events, are of the same magnitude in both the GRIP and the Byrd ice cores. This also rules out the possibility that atmospheric CO₂ concentration changes with similar frequencies and similar durations would be recorded in attenuated form in the Byrd core because of its lower accumulation rate (which causes a broader age distribution of the air enclosed in bubbles at a given depth). However, CO₂ variations with a smaller amplitude parallel to Dansgaard–Oeschger events cannot be excluded with the available results. For example, there is a direct correlation of $r = 0.55$ between CO₂ concentrations and $\delta^{18}\text{O}$ (individual Dansgaard–Oeschger events) for the time interval 26–45 kyr BP, including H3–H5. On the other hand, we suggest, based on Fig. 2d, that atmospheric CO₂ variations and Bond cooling cycles between successive Heinrich events are linked slightly better; for the same time interval of 26–45 kyr BP (two Bond cycles) the correlation between CO₂ and Bond cycles (saw-tooth in Fig. 2a) is 0.57. The higher values of CO₂ correlate quite well with the two Bond cycles following the principal events H5 and H4. These two cycles are also distinct in that their initial mild phases are pronounced and last for more than 2 kyr (following H4) and more than 3 kyr (following H5) rather than just 1 kyr typical for all other Dansgaard–Oeschger events. In contrast, the Bond cycle starting after H3 contains only two very short Dansgaard–Oeschger events that are similar to the later events around 31 and 38 kyr BP. None of these Dansgaard–Oeschger events is reflected in clear atmospheric CO₂ variations, suggesting that they are too short, or too localized in space, to have a net effect on the atmospheric CO₂ concentration. The cycle following H2 is influenced already by the transition from the glacial to the Holocene epoch.

Before discussing the information that this signal carries about the coupled climate–carbon cycle system, the possibility of *in situ* production of CO₂ must be addressed. The impurities of Antarctic ice of the Holocene are generally very low, and carbonate concentrations were found²¹ to be below the detection limit of 0.1 $\mu\text{mol l}^{-1}$. For the glacial period the situation is less clear. Byrd glacial ice is in many respects comparable to GRIP Holocene ice (acidity²², dust concentration⁵, carbonate level²¹). In the latter core, an acid-carbonate reaction or oxidation of organic material takes place, causing a scatter of CO₂ measurements from neighbouring samples ($\sigma \approx 15$ p.p.m.v.) that is much larger than the analytical uncertainty²³ ($\approx \pm 3$ p.p.m.v.). We conclude from the lower scatter in the CO₂ measurements from Byrd glacial ice ($\sigma \approx 4$ p.p.m.v.) that chemical reactions are less important in the Byrd ice core, and that the Byrd CO₂ record probably represents variations of the atmospheric CO₂ concentration. The Vostok (Antarctica) CO₂ record²⁴ is in agreement with the Byrd results⁴, but lacks sufficient resolution for comparison with the present data. Byrd data from the early Holocene agree in magnitude with detailed measurements from D47 (Antarctica) in a short overlapping period (7–6 kyr BP)²⁵. However, Byrd Holocene data show more scatter, due, we believe, to bad core quality in the corresponding depth interval. Detailed CO₂ records should become available from analyses of the already recovered ice cores from Taylor Dome (by the United States) and Dome Fuji (by Japan) and from the cores to be recovered by EPICA (European Project for Ice Coring in Antarctica) at Dome Concordia and Dronning Maud Land. We hope that these records will provide an extension of the investigations to older ages and additional cycles and so confirm the present results. However, a correlation alone does not necessarily exclude the possibility of an artefact. Greenland CO₂ records from Dye 3, Camp Century and Summit all show CO₂ variations of the order of 50 p.p.m.v. parallel to Dansgaard–Oeschger events, although none of these variations are atmospheric. Only a combination of detailed CO₂ and chemical analyses of the various cores will finally allow us decide whether the CO₂ record shown in Fig. 2 indeed represents atmospheric CO₂ variations.

Although sparse, the $\delta^{13}\text{C}$ record of CO₂ (ref. 26) carries some information about the origin of the CO₂ concentration variations.

$\delta^{13}\text{C}$ seems to be rather in anticorrelation than in correlation with the CO₂ concentration. The assumption of an anticorrelation is supported by measurements on an ice core from the South Yamato mountains (Antarctica)²⁷. The *N. pachyderma* record from the North Atlantic indicates that the long-term SST and CO₂ concentration are fairly well correlated. Lower SST correspond to lower CO₂ concentrations. Indeed, a lowered surface temperature leads to an increased CO₂ uptake due to the higher solubility but this leads as well to a lowered $\delta^{13}\text{C}$ value. Because this is not observed in $\delta^{13}\text{C}$, CO₂ uptake by lowered SST is probably not the main reason for the lowering of the CO₂ concentration. Increased activity of the biological pump or increased terrestrial biomass lead to a depletion of the atmospheric CO₂ content parallel to an increase in the $\delta^{13}\text{C}$ ratio²⁸. They are therefore both candidates for explaining the observed CO₂ and $\delta^{13}\text{C}$ variations. The biological pump is the more likely, because one would not expect a higher terrestrial biomass during cold phases. Additional $\delta^{13}\text{C}$ measurements and simulations with dynamical climate models including the carbon cycle²⁹ will be needed to quantify the contributions of the different mechanisms.

The greatest concentrations of atmospheric CO₂ are generally associated with the strongest (in amplitude and duration) Dansgaard–Oeschger cycles following a Heinrich event. Such cycles also represent fast and abrupt changes in the Northern Hemisphere temperature, but there seem to be no significant CO₂ increases parallel to short Dansgaard–Oeschger cycles. We suggest that either the dynamics of these cycles are distinctly different from that of Heinrich events or that the response time for CO₂ to reach a new equilibrium is too long compared to the timescale of these Dansgaard–Oeschger cycles.

Determination of the relation between climate and global CO₂ concentration remains a challenge. Further high-precision $\delta^{13}\text{C}$ and CO₂ measurements, along with climate proxies on a synchronized timescale, are needed to understand the past CO₂ cycle. Such measurements would also be useful in the prediction of future climate change. □

Received 17 February; accepted 24 November 1997.

1. Raynaud, D. *et al.* The ice record of greenhouse gases. *Science* **259**, 926–933 (1993).
2. Oeschger, H. *et al.* in *Climate Processes and Climate Sensitivity* (eds Hansen, J. E. & Takahashi, T.) 299–306 (Vol. 29, Geophys. Monogr. Ser., Am. Geophys. Union, Washington DC, 1984).
3. Stauffer, B., Hofer, H., Oeschger, H., Schwander, J. & Siegenthaler, U. Atmospheric CO₂ concentration during the last glaciation. *Ann. Glaciol.* **5**, 160–164 (1984).
4. Neftel, A., Oeschger, H., Staffelbach, T. & Stauffer, B. CO₂ record in the Byrd ice core 50,000–5,000 years BP. *Nature* **331**, 609–611 (1988).
5. Thompson, L. G. in *Proc. Isotopes and Impurities in Snow and Ice*, Grenoble Aug./Sept. 1975, 351–363 (Publication 118, Int. Assoc. Hydrological Sciences, IAHS-AISH, 1975).
6. Delmas, R. A. A natural artefact in Greenland ice-core CO₂ measurements. *Tellus B* **45**, 391–396 (1993).
7. Ankin, M. *et al.* CO₂ record between 40 and 8 kyr BP from the GRIP ice core. *J. Geophys. Res.* **102**, 26539–26545 (1997).
8. Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).
9. Bond, G. C. & Lotti, R. Iceberg discharges into the North Atlantic on millennial time scales during the last deglaciation. *Science* **267**, 1005–1010 (1995).
10. Maier-Reimer, E. & Mikolajewicz, U. *Experiments with an OGCM on the Cause of the Younger Dryas* (Tech. Rep. 39, Max-Planck-Inst. für Meteorol., Hamburg, 1989).
11. Stocker, T. F. & Wright, D. G. Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes. *Nature* **351**, 729–732 (1991).
12. Wright, D. G. & Stocker, T. F. in *Ice in the Climate System* (ed. Peltier, W. R.) 395–416 (NATO ASI Ser. I, 12, Springer, Berlin, 1993).
13. Oeschger, H., Neftel, A., Staffelbach, T. & Stauffer, B. The dilemma of the rapid variations in CO₂ in Greenland ice cores. *Ann. Glaciol.* **10**, 215–216 (1988).
14. Sowers, T. & Bender, M. Climate records covering the last deglaciation. *Science* **269**, 210–214 (1995).
15. Blunier, T. *et al.* Timing of the Antarctic Cold Reversal and the atmospheric CO₂ increase with respect to the Younger Dryas event. *Geophys. Res. Lett.* **24**, 2683–2686 (1997).
16. Chappellaz, J. *et al.* Changes in the atmospheric CH₄ gradient between Greenland and Antarctica during the Holocene. *J. Geophys. Res.* **102**, 15987–15999 (1997).
17. Chappellaz, J. *et al.* Synchronous changes in atmospheric CH₄ and Greenland climate between 40 and 8 kyr BP. *Nature* **366**, 443–445 (1993).
18. Johnsen, S. J., Dahl-Jensen, D., Dansgaard, W. & Gundestrup, N. Greenland paleotemperatures derived from GRIP bore hole temperature and ice core isotope profiles. *Tellus B* **47**, 624–629 (1995).
19. Schwander, J. *et al.* Age scale of the air in the Summit ice: Implication for glacial-interglacial temperature change. *J. Geophys. Res.* **102**, 19483–19494 (1997).
20. Bond, G. *et al.* Correlations between climate records from North Atlantic sediments and Greenland ice. *Nature* **365**, 143–147 (1993).
21. Fuhrer, A. *Ein System zur Messung des totalen Kohlenstoffgehaltes polarer Eisproben* Thesis, Univ. Bern (1995).
22. Hammer, C. U., Clausen, H. B. & Langway, C. C. Jr Electrical conductivity method (ECM) stratigraphic dating of the Byrd Station ice core, Antarctica. *Ann. Glaciol.* **20**, 115–120 (1994).

23. Anklin, M. J. *Kohlenstoffdioxid Bestimmungen in Luftproben aus einem neuen Tiefbohrkern von Summit (Grönland)* Thesis, Univ. Bern (1994).
24. Barnola, J.-M., Pimienta, P., Raynaud, D. & Korotkevich, Y. S. CO₂-climate relationship as deduced from the Vostok ice core: A re-examination based on new measurements and on a re-evaluation of the air dating. *Tellus* **43**, 83–90 (1991).
25. Barnola, J.-M., Jeanjean, E. & Raynaud, D. Holocene atmospheric CO₂ evolution as deduced from an Antarctic ice core. *Eos* **77**, 151 (1996).
26. Leuenberger, M., Siegenthaler, U. & Langway, C. C. Carbon isotope composition of atmospheric CO₂ during the last ice age from an Antarctic ice core. *Nature* **357**, 488–490 (1992).
27. Machida, T. *et al.* Variations of the CO₂, CH₄ and N₂O concentrations and $\delta^{13}\text{C}$ of CO₂ in the glacial period deduced from an Antarctic ice core, south Yamato. *Proc. NIPR Symp. Polar Meteorol. Glaciol.* **10**, 55–65 (1996).
28. Keir, R. S. On the Late Pleistocene ocean geochemistry and circulation. *Paleoceanography* **3**, 413–445 (1988).
29. Marchal, O., Stocker, T. F. & Joos, F. Impact of oceanic reorganisations and the marine carbon cycle and atmospheric CO₂. *Paleoceanography* (submitted).

Acknowledgements. We thank M. Leuenberger and J.-M. Barnola for discussions, and F. Finet for measuring additional CH₄ samples at LGGE. This work, as part of the Greenland Ice Core Project (GRIP), was supported by the University of Bern, the Swiss National Science Foundation, the Federal Department of Energy (BEW), the Schwerpunktprogramm Umwelt (SPPU) of the Swiss National Science Foundation, the EC programme "Environment and Climate 1994–1998", the Fondation de France and the Programme National de Dynamique du Climat de CNRS.

Correspondence and requests for materials should be addressed to B.S. (e-mail: stauffer@climate.unibe.ch).

Influence of sea-salt on aerosol radiative properties in the Southern Ocean marine boundary layer

D. M. Murphy*, J. R. Anderson†, P. K. Quinn‡, L. M. McInnes§, F. J. Brechtel||, S. M. Kreidenweis||, A. M. Middlebrook*, M. Pósfai#, D. S. Thomson* & P. R. Buseck†#

* Aeronomy Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado 80303, USA

† Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287, USA

‡ Pacific Marine Environment Laboratory, National Oceanic and Atmospheric Administration, Seattle, Washington 98115, USA

§ Climate Monitoring and Diagnostics Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado 80303, USA

|| Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado 80523, USA

¶ Cooperative Institute for Research in the Environmental Sciences, University of Colorado, Boulder, Colorado 80309, USA

Department of Geology, Arizona State University, Tempe, Arizona 85287, USA

There has been considerable debate about the relative importance of sea-salt and sulphate from non-sea-salt sources in determining aerosol radiative effects in the marine boundary layer. In the marine boundary layer, the most numerous aerosols are volatile sulphate particles smaller than about 0.08 μm (ref. 1) and most of the aerosol mass is in a few sea-salt particles larger than 1 μm . Yet intermediate-size aerosols between about 0.08 and 1 μm diameter are the most relevant to the radiative forcing of climate because they efficiently scatter solar radiation and also serve as cloud nuclei². Indeed, Charlson *et al.*³ hypothesized that oceanic production of sulphate aerosols from the oxidation of dimethyl sulphide could be a powerful feedback in the climate system. It is generally assumed that marine aerosols smaller than about 1 μm are non-sea-salt sulphate, but a recent review cites indirect evidence that many aerosols in the sub-micrometre range contain at least some sea-salt^{4,5}. Here we present direct observational evidence from a remote Southern Ocean region that almost all aerosols larger than 0.13 μm in the marine boundary layer contained sea-salt. These sea-salt aerosols had important radiative

effects: they were responsible for the majority of aerosol-scattered light, and comprised a significant fraction of the inferred cloud nuclei.

The first Aerosol Characterization Experiment (ACE-1) included complementary techniques to assess the chemical and radiative properties of aerosols in the marine boundary layer (MBL). The composition of aerosols was measured by electron microscopy and mass spectrometry of single particles, aerosol volatility and bulk analysis of size-segregated aerosol samples. At Cape Grim, Tasmania (41°S, 145°E), the elemental composition of individual particles collected on filters and grids was determined using automated scanning and manual transmission electron microscopy (SEM),

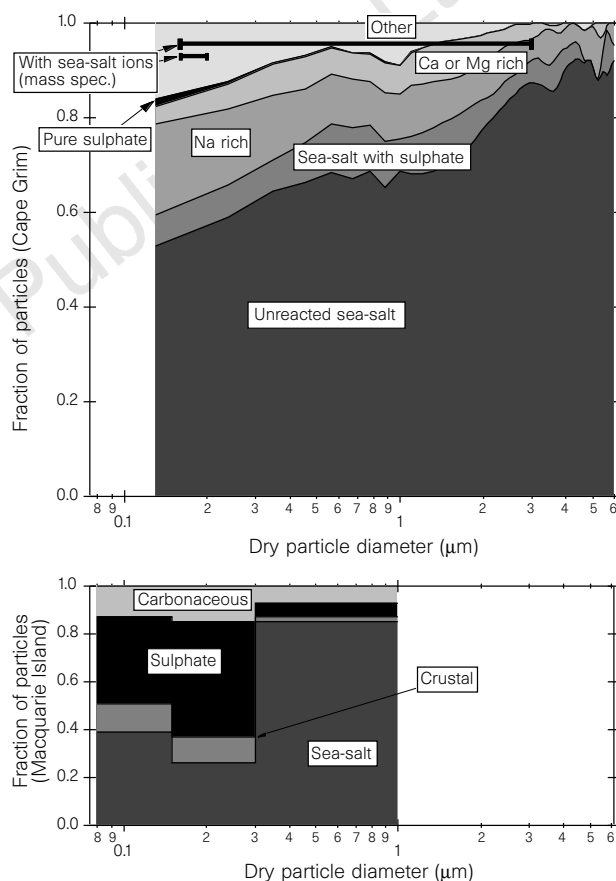


Figure 1 The chemical composition of particles at Cape Grim and Macquarie Island as a function of their diameter. Top, composition at Cape Grim from automated SEM analysis (filled areas, $n = 19,640$). Also shown are the fraction of particles analysed by mass spectrometry that contained detectable sea-salt ions ($n = 5,240$). The two bars are for samples taken with or without a mobility analyser to select only a limited size range for analysis. Bottom, composition at Macquarie Island from manual TEM analysis ($n = 4$ samples, 300 particles). SEM data at Cape Grim were classified using cluster analysis, then the clusters were grouped into the categories shown here 'Unreacted sea-salt' includes particles with elemental compositions very close to natural sea-salt. The 'other' category includes particles that contained only light elements, salt particles with compositions far from sea-salt, and unusual particles such as soot or metals. For samples from Macquarie Island, sea-salt-containing particles were defined as those containing appropriate quantities of Na, Mg, Ca, K and variable quantities of S, with or without measurable quantities of Cl. In the mass spectra, particles containing sea-salt were identified by the presence of Cl in the negative ions and Na or K in the positive ions. The mass spectrometer measured a higher fraction of sea-salt-containing particles because it is very sensitive to Cl and Na. However, as typified in Fig. 2, most of the mass spectra showed a substantial fraction of sea-salt.

TEM) with energy dispersive X-ray analysis^{6,7}. Electron microscope samples were taken only during clean air conditions from the 55 m level of a tower located on a 90 m cliff next to the ocean. Single particles sampled at Cape Grim were also analysed *in situ* with a laser ionization mass spectrometer^{8,9}. The elemental composition of individual particles collected at Macquarie Island (54° S, 158° E) was independently determined using TEM with energy dispersive X-ray analysis^{10–12} and aerosol volatility. Onboard the NOAA ship *Discoverer*, a 7-stage multi-jet Berner-type impactor was used to sample aerosol from an inlet 18 m above the ocean surface for gravimetric and ion chromatographic analysis¹³.

The measurements made at Cape Grim, Macquarie Island, and onboard the *Discoverer* all showed sea-salt in the accumulation mode. Both the SEM and mass spectrometer data show that many particles at Cape Grim, even those smaller than 0.2 µm, contained elements characteristic of sea-salt (Figs 1 and 2). During unpolluted conditions, over 90% of aerosol particles larger than 0.13 µm diameter contained sea-salt and fewer than 1% of these were pure sulphate particles. The percentage of particles containing sea-salt dropped rapidly for smaller particles, with some suggestion (Fig. 1) that the change from sea-salt to sulphate was at a slightly larger size at Macquarie Island than at Cape Grim. For particles between 0.05 and 0.15 µm diameter, independent manual TEM analyses found sea-salt in 5 to 25% of the particles at Cape Grim and 5 to 47% of particles at Macquarie Island. Most 0.05 µm diameter aerosols were volatile at temperatures between 95 and 150 °C, indicating that they did not contain sea-salt. Microdiffraction patterns of particles smaller than 0.1 µm suggested an ammonium sulphate crystalline structure.

These data, which indicate a concentration of sea-salt-containing particles from about 30 to over 100 per cm³, differ from some previous estimates of fewer than 1 sea-salt particle per cm³ in the MBL^{3,14}. However, they agree with a recent parameterization of sea-salt particles⁴. The average wind speed at Cape Grim during clean periods of ACE-1 was 11 m s⁻¹. High wind speeds assist in the production of new sea-salt particles^{4,15}. However, it is difficult to explain the dominant role for sea-salt in ACE-1 based on wind speed alone. Some previous measurements in the Northern Hemisphere may have been influenced by anthropogenic sulphate, resulting in a larger role for sulphate particles than in the ACE-1 study area^{16,17}. For example, SEM analysis of particles at Spitsbergen and Bermuda by the same techniques used at Cape Grim found large concentrations of sulphate particles larger than 0.1 µm^{7,18}. Measurements elsewhere in the MBL have shown mixed results on the relative

importance of sulphate and sea-salt to submicrometre particles^{19–22}. The ACE-1 study area had low sulphate concentrations: measurements on the *Discoverer* showed 150 ng m⁻³ compared to 1,300 ng m⁻³ for the Northern Hemisphere mid-latitudes and 400 ng m⁻³ for the equatorial Pacific²³.

The fraction of aerosol scattering and backscattering due to sea-salt or non-sea-salt (n.s.s.) sulphate aerosol was calculated using the aerosol chemical composition from impactor measurements, the particle size distribution and Mie scattering theory¹³. Size-segregated aerosol samples collected onboard the *Discoverer* were analysed gravimetrically and with ion chromatography (Fig. 3a). The ionic mass consisted of sea-salt, n.s.s. sulphate plus associated ammonium, and methanesulphonate (MSA, an oxidation product of dimethylsulphoxide, DMS). The size distribution for the calculations was derived from a differential mobility analyser for particles 0.02 to 0.57 µm in diameter and an aerodynamic sizer for particles from 0.85 to 5 µm. Measured sizes were converted to geometric sizes using densities derived from the impactor data¹³. Particles were assumed to be spherical and the sea-salt and n.s.s. sulphate were assumed to be in different particles. However, the assumption of an external or internal mixture does not significantly affect scattering calculations of salt and sulphates²⁴.

Most of the light scattering by the aerosol was due to sea-salt (Fig. 3b). Even considering just the submicrometre portion of the aerosol, over 75% of either the total scattering or backscattering was due to sea-salt. Sea-salt appears to have been the dominant chemical

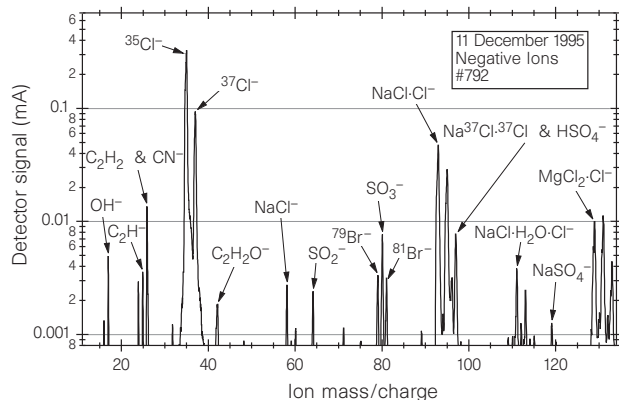


Figure 2 A typical negative ion mass spectrum of a single particle at Cape Grim. This particular spectrum is chosen to be as close as possible to the average of 0.17 µm dry diameter particles, as determined by a differential mobility analyser. The dominant peaks are from sea-salt, but sulphate and organic peaks are also present.

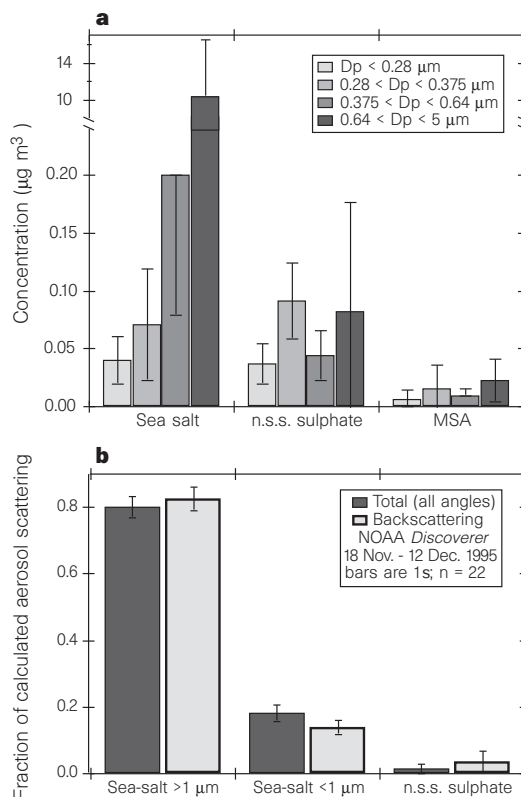


Figure 3 Composition and backscattering of aerosols. **a**, The chemical composition of aerosols for several size ranges measured from the NOAA *Discoverer* in the Southern Ocean. **b**, The calculated light scattering and backscattering at 550 nm due to sub- and supermicrometre sea-salt and n.s.s. sulphate. Because of its small mass fraction, the scattering due to MSA was negligible. Sea-salt dominated both the total scattering (all angles) and backscattering (hemispheric) by the aerosol. Total aerosol calculated backscatter was between 0.0005 and 0.004 km⁻¹.

constituent in determining the direct radiative effects of the aerosol in this remote Southern Ocean region.

The radiative forcing of aerosols through their influence on cloud nucleation processes has been estimated to be as large as that of the direct effect but has a much larger associated uncertainty which is partly due to the complex relationship between aerosol concentration and cloud droplet concentration²⁵. Changes in the number concentration of cloud condensation nuclei (CCN) can have an especially great influence on cloud albedo in the cleanest environments. In the MBL, the approximate size of particles serving as the smallest cloud nuclei can be determined from a minimum in the aerosol size distribution²⁶. Particles that act as cloud nuclei acquire additional mass nuclei, because of both liquid-phase chemistry and the large surface area for uptake of condensable species. When the clouds evaporate, the nuclei are larger than they were before the cloud processing. Smaller particles that do not act as cloud nuclei do not experience this growth. Repeated cycling through clouds creates a minimum in the size distribution at diameters of about 0.08 μm (Fig. 4a). We infer the number of cloud nuclei as $N_{0.08}$, the number of particles larger than 0.08 μm . This definition has the advantage that it reflects the supersaturations actually achieved in the MBL.

From the discussion above of sea-salt content in submicrometre aerosols, it might be expected that a large fraction of $N_{0.08}$ in the Southern Ocean region contained sea-salt. Measurements from Macquarie Island confirm this. The concentrations of non-volatile particles at 300 °C (Figure 4b) usually accounted for more than 50% of $N_{0.08}$. The time series of non-volatile particles and $N_{0.08}$ were also

correlated. In addition to measuring the number of non-volatile particles at various temperatures, subsets of non-volatile particles were collected with a low-pressure impactor ($D_{50} = 0.06 \mu\text{m}$) for TEM analysis. These samples confirmed that most of the non-volatile particles at 300 °C were sea-salt and were large enough to act as cloud nuclei.

The correlation between seasonal cycles of CCN and MSA at Cape Grim has been used to support the theory that sulphate influences marine CCN concentrations^{27,28}. However, the strongest correlations were for CCN measured at a 1.2% supersaturation, which activates particles much smaller than those that actually serve as cloud nuclei in the MBL. Such smaller particles are mostly sulphate. The correlations were weaker at 0.23% supersaturation, which is more representative of the MBL¹. In addition, most of our data were taken before the DMS production in the ACE-1 region reached its annual peak in December of 1995.

Our results indicate an important role for submicrometre sea-salt particles in determining the radiative properties of aerosols in the remote MBL. The SEM, TEM, mass spectrometer and bulk analyses are all consistent with the presence of sea-salt in the vast majority of particles larger than 0.13 μm dry diameter. Although most of these particles contained sea-salt, the radiative properties are also influenced by other species. Non-seasalt sulphate may contribute enough mass to enlarge some of the smallest sea-salt particles to more effective cloud nuclei. The Cape Grim data show that many aerosols were internal mixtures (in the same particles) of sea-salt and sulphates. In such mixtures it is difficult to separate the contributions of sulphate and salt, especially because salt aerosols provide important reaction sites for converting SO_2 to sulphate^{17,29}. The Cape Grim mass spectrometer data indicate that organics also contribute a small fraction of the content of the submicrometre salt particles³⁰.

If sulphate particles acted as cloud nuclei, then some of them should have acquired enough additional mass through cloud processing to be large enough to be observed by the SEM and mass spectrometer. The rarity of pure sulphate particles larger than 0.13 μm at Cape Grim shows that particles formed through the condensation of sulphur-containing precursor gases did not often grow that large, at least not without coagulating with some sea-salt. This places a strong constraint on models of the dynamics of aerosols in the marine boundary layer, and may even indicate that few pure sulphate particles were acting as cloud nuclei in that region.

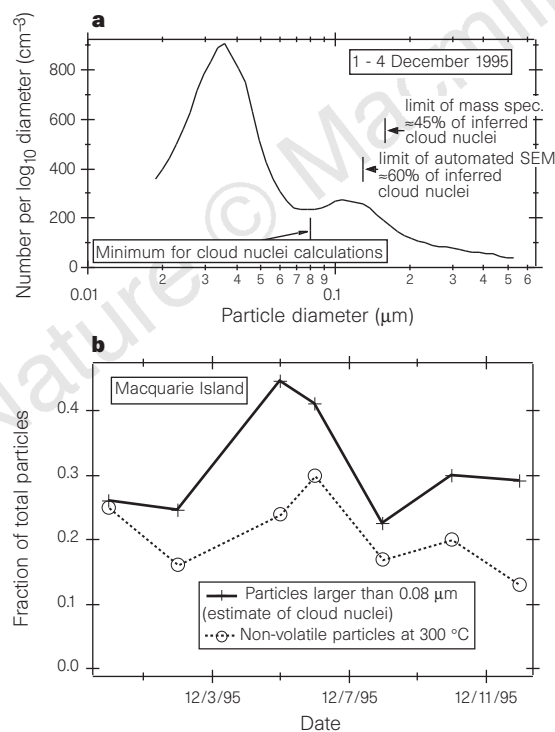


Figure 4 Size distribution and number of particles. **a**, An average size distribution from Macquarie Island while measuring pristine marine air. The distribution shows the characteristic dip near 0.08 μm caused by growth of certain particles in clouds²⁶. Most larger particles are cloud nuclei and most smaller particles are not. The units on the vertical axis are chosen for visual comparison: the number of particles in a given size range is proportional to the area under the curve in that range. The integral of this distribution is 470 cm^{-3} , which is a typical concentration of particles in a remote marine environment. **b**, The number of non-volatile particles and number of inferred cloud nuclei as fractions of the total number of particles measured at 25 °C. (Date is given as month/day/year.)

Received 3 June; accepted 2 December 1997.

- Hoppel, W. A., Frick, G. M. & Fitzgerald, J. W. Deducing droplet concentration and supersaturation in marine boundary layer clouds from surface aerosol measurements. *J. Geophys. Res.* **101**, 26553–26565 (1996).
- Schwartz, S. E. The whitehouse effect—shortwave radiative forcing of climate by anthropogenic aerosols: an overview. *J. Aerosol. Sci.* **27**, 359–382 (1996).
- Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. Oceanic phytoplankton, atmospheric sulphur, cloud albedo, and climate. *Nature* **326**, 655–661 (1987).
- O'Dowd, C. D. & Smith, M. H. Physicochemical properties of aerosols over the Northeast Atlantic: evidence for wind-speed-related submicron sea salt production. *J. Geophys. Res.* **98**, 1137–1149 (1993).
- O'Dowd, C. D., Smith, M. H., Consterdine, I. E. & Lowe, J. A. Marine aerosol, sea-salt, and the marine sulfur cycle: a short review. *Atmos. Environ.* **31**, 73–80 (1997).
- Pósfai, M., Anderson, J. R., Buseck, P. R. & Sievering, H. Compositional variations of sea-salt-mode aerosol particles from the North Atlantic. *J. Geophys. Res.* **100**, 23063–23074 (1995).
- Anderson, J. R., Buseck, P. R. & Patterson, T. L. Characterization of the Bermuda tropospheric aerosol by combined individual-particle and bulk-aerosol analysis. *Atmos. Environ.* **30**, 319–338 (1996).
- Murphy, D. M. & Thomson, D. S. Chemical composition of single aerosol particles at Idaho Hill: negative ion measurements. *J. Geophys. Res.* **102**, 6353–6368 (1997).
- Murphy, D. M., Thomson, D. S., Middlebrook, A. M. & Schein, M. E. In situ single particle characterization at Cape Grim. *J. Geophys. Res.* ACE-1 special issue (in the press).
- McInnes, L. M., Covert, D. S. & Baker, B. The number of sea-salt, sulfate, and carbonaceous particles in the marine atmosphere: TEM measurements consistent with the ambient size distribution. *Tellus* **49**, 300–313 (1997).
- Brechtl, F., Kreidenweis, S. & Swan, H. B. Air mass characteristics, total particle concentrations, and size distributions at Macquarie Island, Tasmania during ACE-1. *J. Geophys. Res.* ACE-1 special issue (in the press).
- Kreidenweis, S. M., McInnes, L. M. & Brechtl, F. J. Observations of aerosol composition at Macquarie Island during ACE-1. *J. Geophys. Res.* ACE-1 special issue (submitted).
- Quinn, P. K., Kapustin, V. N., Bates, T. S. & Covert, D. S. Aerosol optical properties in the marine boundary layer during ACE-1 and the underlying chemical and physical aerosol properties. *J. Geophys. Res.* (in the press).

14. Hobbs, P. V. Simultaneous airborne measurements of cloud condensation nuclei and sodium-containing particles over the ocean. *Q. J. R. Meteorol. Soc.* **97**, 263–271 (1971).
15. Blanchard, D. C. The oceanic production rate of cloud nuclei. *J. de Rech. Atmos.* **4**, 1–6 (1969).
16. Twomey, S. On the composition of cloud nuclei in the northeastern United States. *J. de Rech. Atmos.* **4**, 281–285 (1968).
17. Huebert, B. J., Zhuang, L., Howell, S., Noone, K. & Noone, B. Sulfate, nitrate, methanesulfonate, chloride, ammonium, and sodium measurements from ship, island, and aircraft during the Atlantic Stratocumulus Transition Experiment/Marine Aerosol Gas Exchange. *J. Geophys. Res.* **101**, 4413–4423 (1996).
18. Anderson, J. R., Buseck, P. R., Saucy, D. A. & Pacyna, J. M. Characterization of individual fine-fraction particles from the arctic aerosol at Spitsbergen, May–June 1987. *Atmos. Environ.* **26A**, 1747–1762 (1992).
19. O'Dowd, C. D. *et al.* Biogenic sulphur emissions and inferred non-sea-salt-sulphate cloud condensation nuclei in and around Antarctica. *J. Geophys. Res.* **102**, 12839–12854 (1997).
20. Clarke, A. D., Ahlquist, N. C. & Covert, D. S. The Pacific marine aerosol: evidence for natural sulfates. *J. Geophys. Res.* **92**, 4179–4190 (1987).
21. Rojas, C. M. & Van Grieken, R. Electron microprobe characterization of individual aerosol particles collected by aircraft above the southern bight of the North Sea. *Atmos. Environ.* **26A**, 1231–1237 (1992).
22. Quinn, P. K., Marshall, S. F., Bates, T. S., Covert, D. S. & Kapustin, V. N. Comparison of measured and calculated aerosol properties relevant to the direct radiative forcing of tropospheric sulfate aerosol on climate. *J. Geophys. Res.* **100**, 8977–8991 (1995).
23. Quinn, P. K., Bates, T. S., Johnson, J. E., Covert, D. S. & Charlson, R. J. Interactions between sulfur and reduced nitrogen cycles over the central Pacific Ocean. *J. Geophys. Res.* **95**, 16405–16416 (1990).
24. Tang, I. Thermodynamic and optical properties of mixed-salt aerosols of atmospheric importance. *J. Geophys. Res.* **102**, 1883–1893 (1997).
25. Fouquart, Y. & Isaka, H. Sulfur emission, CCN, clouds and climate: a review. *Ann. Geophys.* **10**, 462–471 (1992).
26. Hoppel, W. A., Frick, G. M. & Larson, R. E. Effect of nonprecipitating clouds on the aerosol size distribution in the marine boundary layer. *Geophys. Res. Lett.* **13**, 125–128 (1986).
27. Bigg, E. K., Gras, J. L. & Evans, C. Origin of Aitken particles in remote regions of the Southern Hemisphere. *J. Atmos. Chem.* **1**, 203–214 (1984).
28. Ayers, G. P. & Gras, J. L. Seasonal relationship between cloud condensation nuclei and aerosol methanesulphonate in marine air. *Nature* **353**, 834–835 (1991).
29. Sievering, H. *et al.* Ozone oxidation of sulfur in sea-salt aerosol particles during the Azores Marine Aerosol and Gas Exchange experiment. *J. Geophys. Res.* **100**, 23075–23081 (1995).
30. Middlebrook, A. M., Murphy, D. M. & Thomson, D. S. Observations of organic material in individual marine particles at Cape Grim during ACE-1. *J. Geophys. Res.* ACE-1 special issue (in the press).

Acknowledgements. This research is a contribution to the International Global Atmospheric Chemistry (IGAC) Core project of the International Geosphere-Biosphere Programme (IGBP) and is part of the IGAC Aerosol Characterization Experiments (ACE). The assistance of the Cape Grim Baseline Air Pollution Station staff and the Australian Bureau of Meteorology and the Commonwealth Scientific and Industrial Research Organization is acknowledged. Work at Macquarie Island was sponsored by the Office of Naval Research and electron microscope work at Cape Grim was supported by the NSF Atmospheric Chemistry Program. Work aboard the *Discoverer* was funded by the Aerosols Project of NOAA's Climate and Global Change Program. F.J.B. acknowledges an EPA fellowship and A.M.M. acknowledges a DOE fellowship.

Correspondence and requests for materials should be addressed to D.M.M. (e-mail: murphyd@al.noaa.gov).

Transition from dome-forming to plinian eruptive styles controlled by H₂O and Cl degassing

Benoît Villemant*† & Georges Boudon‡†

* LGCS, CNRS URA 1758 et Université P.M. Curie, boîte 109, 4 place Jussieu, 75252 Paris Cedex 05, France

† Département de Volcanologie, Institut de Physique du Globe de Paris, 4 place Jussieu, 75252 Paris cedex 05, France

‡ Laboratoire de Géomatériaux, CNRS URA 734 et Observatoires Volcanologiques, Institut de Physique du Globe de Paris, boîte 89, 4 place Jussieu, 75252 Paris Cedex 05, France

The transition from a plinian (pumice) to an effusive (dome-forming) eruptive style is frequently observed in volcanic systems and is generally attributed to the progressive loss of volatiles from magma stored in a superficial reservoir. This explosive–effusive transition has been explained by the evolution from a closed to an open system of degassing^{1–4}. But in this context, an eruption at Mt Pelée (Martinique, French West Indies) dated at 650 years ago, which exhibited a rarely observed^{5,6} succession from dome-forming to plinian activity in a short interval of time⁷, is at odds with such an explanation. In this eruption, near-surface explosions of the dome produced two peléean turbulent pyroclastic flows, whose

deposits are similar to those of the effusive 1902 eruption, and then plinian activity produced pumice fallouts and flows. The reconstruction of the degassing paths of both eruptive regimes using the densities and the H₂O and Cl contents of the clasts shows that the interaction of rising magma with hydrothermal fluids at shallow depth may play a critical role in determining eruptive style.

The differences in eruptive styles during the Mt Pelée eruption at 650 years before present (650 yr BP; the P1 eruption) are not related to differences in pre-eruptive conditions, as shown by the homogeneity of the mineralogical and the trace- and major-element compositions of all erupted magmas^{8,9}. However, the existence of ²³⁸U–²³⁰Th disequilibrium and some mobile-element anomalies (for example, arsenic) show that the dome-forming magma chemically interacted with fluids from the near-surface hydrothermal system⁸. Magma erupted during the plinian activity is in radioactive equilibrium and chemically homogeneous, showing that it has remained as a closed system from the magma source. To assess the degassing history of both eruptive regimes, geochemical and textural measurements have been made on clasts selected from the

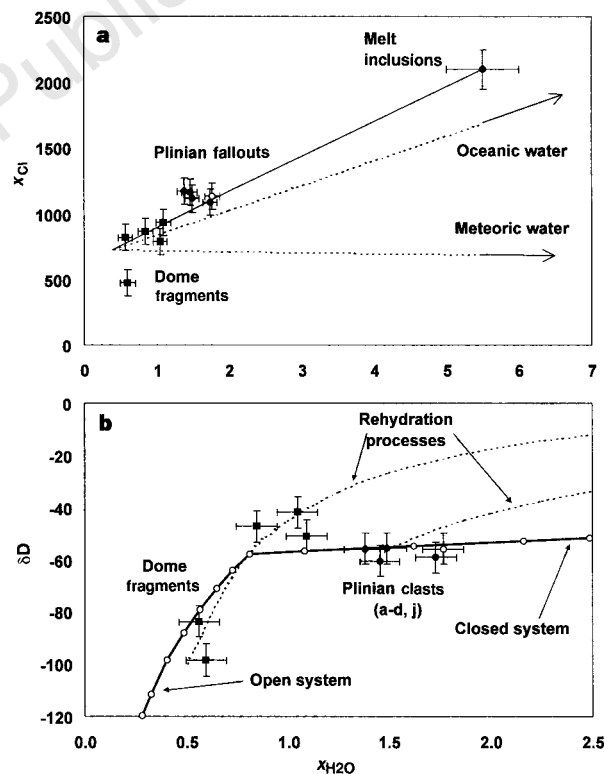


Figure 1 Correlations between $x_{\text{H}_2\text{O}}$, x_{Cl} and δD . Trends are shown for volcanic samples and also for rehydration by meteoric or oceanic water (dashed lines). **a**, H₂O and Cl contents of all clasts and pre-eruptive melts are well correlated, indicating a close degassing behaviour for both species. **b**, δD and H₂O contents may be well explained by simple closed- (plinian clasts) and open- (dome fragments) system degassing. (x_{Cl} is in p.p.m., $x_{\text{H}_2\text{O}}$ is in %.) Closed-system degassing model is calculated for an initial δD value of -40 and $\alpha_{\text{vapour-melt}} = 1.02$. Open-system degassing model is calculated using a Rayleigh distillation law with an initial δD value of -58 and $\alpha_{\text{vapour-melt}} = 1.06$. Rehydration models (dashed lines) have been calculated using mixing with pure water with $\delta D = +10\text{‰}$. We note that late contamination cannot be ruled out for dome products, but the initial chemical homogeneity of P1 eruption magmas requires an open-system degassing to produce the less-H₂O-rich dome fragments. δD values are from B. R. Rahaingoson and P. A. Agrinier (personal communication). Here $\delta D = \{[(D/H)_R/(D/H)_S] - 1\} \times 10^3$, where R and S stand for rock and standard, respectively; $\alpha_{\text{vapour-melt}} = (D/H)_{\text{vapour}}/(D/H)_{\text{melt}}$.

different eruptive units⁹. The data reported here are part of this study.

Densities of clasts, crystallinities, and H₂O and Cl contents measured on individual clasts from the two peléean ash-flow deposits (dome fragments) and from two plinian pumice layers (a–d and j) are given in Table 1. Following the procedure given in ref. 10, the volume ratio of gas to melt (V_g/V_l) are estimated by the volume ratio of vesicles to groundmass. H₂O and Cl weight fractions in the melts (x_{H_2O} and x_{Cl}) are calculated by correcting bulk rock contents for crystallinities (weight fraction of phenocrysts). These values are consistent with direct measurements in glasses (see Table 1). Pre-eruptive H₂O (ref. 11) and Cl contents are obtained from melt inclusion analyses. The measured bulk-rock H₂O contents are in the upper range reported for similar eruptive systems¹⁰. Rehydration processes by meteoric water (which are often suspected to have modified glass compositions after eruption) are excluded on the basis of Cl/H₂O and δD measurements. x_{H_2O} and x_{Cl} of clasts and melt inclusions display a clear correlation, with a slope that is distinct from that which would result from contamination of the magmatic volatiles by mixing with meteoric or oceanic water (Fig. 1a). The δD – x_{H_2O} correlations (B. R. Rahaingoson and P. A. Agrinier, personal communication, Fig. 1b) and the Th isotope results⁸ cannot be interpreted in terms of a simple rehydration process, but are consistent with open-system degassing of dome products and closed-system degassing of plinian clasts. Moreover, scanning electron microscopy studies do not show any evidence of glass alteration⁹.

In our degassing model, we start by assuming that degassing of a H₂O-rich magma occurs at equilibrium, the weight fraction of H₂O in the melt (at pressure P) above the saturation depth is given by the Burnham's solubility equation^{12,13}:

$$x_{H_2O} = S_{H_2O} P^{n_{H_2O}} \quad (1)$$

where S_{H_2O} and n_{H_2O} are constants (see Fig. 2). If the gas does not escape from the magma (closed system), V_g/V_l is given by the perfect gas law:

$$V_g/V_l = K_R(x_{H_2O} - x_{ps})/P \quad (2)$$

where K_R is a constant and x_{ps} is the initial water content (P_s is saturation pressure). The estimate of pre-eruptive water content allows us to calculate the theoretical closed-system evolution path in the x_{H_2O} – V_g/V_l diagram (Fig. 2a). H₂O saturation occurs at around 200 MPa (~8 km depth) for the water content required by melt inclusion and phase equilibria¹¹. We assume a very simplified thermal model in which temperature is constant until some quenching event at which temperature is suddenly reduced below the glass transition temperature. The pre-eruptive Cl content of the magma, as measured in melt inclusions (Table 1) is close to the Cl solubility limit at 200 MPa in rhyolitic H₂O-saturated melts^{14–16}. The weight fraction of Cl in the melt is therefore controlled by the H₂O vapour exsolution:

$$x_{Cl} = x_{Cl}^0/[1 + d_{v-1}^Cl(x_{H_2O}^0 - x_{H_2O})] \quad (3)$$

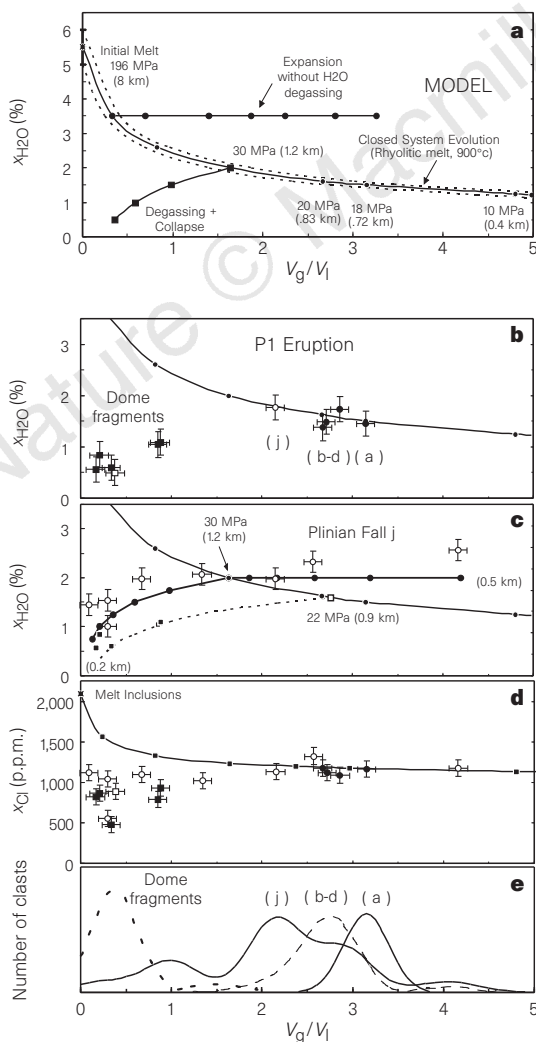


Figure 2 Comparison of model results with measured rock compositions. **a**, Theoretical evolution (evolution with no gas escape from the magma column) is calculated using the solubility law (equation (1)) and the perfect-gas law (equation (2)), and assuming an initial H₂O content of $5.5 \pm 0.5\%$ (ref. 11). The values of S_{H_2O} ($=0.038$) and n_{H_2O} ($=0.54$) in equation (1) have been calculated for P1 eruption melt composition (rhyolitic; $SiO_2 \approx 69.9 \pm 2.5\%$) and temperature ($T \approx 900^\circ C$; ref. 11) using Burnham's equations^{12,13}. In equation (2), $K_R = (R/M_{H_2O})T\rho_l$, where R represents the perfect gas constant, M_{H_2O} the molar mass of H₂O and ρ_l the magma density. T is assumed constant and equal to $900^\circ C$. $K_R = 14.0$ if P is expressed in MPa. Confining gas pressure of erupting magma is higher or equal to the lithostatic load and thus may be converted in minimum depth (expressed in kilometres under the summit of the volcano and calculated using a mean lithostatic load of $2.5 g cm^{-3}$). 'Expansion without H₂O degassing' is calculated assuming that H₂O diffusion in the melt becomes negligible after fragmentation, whereas melt viscosity remains low enough to allow continuous vesicle expansion with decreasing pressure (see also ref. 10). 'Degassing + collapse' shows the calculated degassing of a permeable melt, controlled by H₂O solubility and vesicle collapse kinetics (see text). **b**, Data for the P1 eruption. Solid squares, dome fragments; dots, clasts from the first plinian fallout (a–d); open circles, clast from the main density mode of plinian fallout (j). V_g/V_l distribution of each unit is given in **e**. **c**, H₂O degassing model for the whole plinian fallout (j) and dome products. Solid line and dots, best fit for layer (j) samples. Two distinct evolution paths may be distinguished from a common 'closed-system' initial value (30 MPa, 1.2 km): dense clasts ($V_g/V_l < 1.5$, ~20% of the population) correspond to the degassing of a permeable magma; vesiculated clasts ($V_g/V_l > 3$, ~10% of the population) correspond to vesicle expansion in closed system without H₂O diffusion. Dotted line corresponds to an open system degassing model for dome fragments (small squares) deduced from the above model; initial closed system value (open square) corresponds to a pressure of ~22 MPa and a depth of ~0.9 km. **d**, Cl degassing. The line corresponds to the model of equilibrium partitioning between melt and H₂O vapour phase ($d_{v-1}^{Cl} = 20$ is assumed^{15,16}). Cl initial content is measured in melt inclusions (see Table 1). The similarity of Cl and H₂O behaviour is also demonstrated by the correlation in Fig. 1a. **e**, Distribution of V_g/V_l ratio in the three studied units. Vertical axis, number of clasts; arbitrary units differing for each population (dome fragments, 203 samples; unit a, 137 samples; unit b–d, 281 samples; unit j, 91 samples). Densities corresponding to the main distribution modes of V_g/V_l volume ratio (d is calculated on a crystal-free basis): plinian fallout (j); $d_1 = 0.51$, $d_2 = 0.66$, $d_3 = 1.29$; plinian fallout (a–d), $d = 0.68 \pm 0.07$; dome fragments, 2.20 ± 0.20 .

Table 1 Analyses of erupted material and initial melt

		Dome						First pumice fallout				Pumice fallout j								Initial melt
Sample	(1)	1311E	1311F1	1311F3	901F1	901F4	V7-M	1310A1	1310B1	1310C1	1310D1	J5-8	J5-5	1310J1	J5-31	J5-60	J5-69	1215J1	J5-70	
Crystallinity	(2)	0.38	0.29	0.27	0.29	0.33	0.20	0.24	0.27	0.26	0.25	0.34	0.35	0.32	0.30	0.26	0.36	0.39	0.30	
Density	(3)	1.40	2.15	1.38	2.22	1.94	1.87	0.62	0.70	0.70	0.67	0.50	0.73	0.82	1.11	1.55	1.99	2.00	2.37	2.59 ± 0.03
Vesicularity (%)	(4)	46	17	47	14	25	28	76	73	73	74	81	72	68	57	40	23	22	8	
V_g/V_l	(5)	0.85	0.20	0.88	0.16	0.33	0.38	3.15	2.67	2.72	2.86	4.16	2.57	2.15	1.34	0.67	0.30	0.29	0.09	
H ₂ O (%)	(6)	0.65	0.60	0.80	0.40	0.40	0.40	1.10	1.00	1.10	1.30	1.60	1.40	1.20	1.30	1.30	0.80	0.40	0.80	5.5 ± 0.5 (7)
x _{H₂O}	(8)	1.05	0.85	1.09	0.56	0.60	0.50	1.45	1.38	1.49	1.73	2.43	2.16	1.77	1.86	1.75	1.25	0.65	1.15	
Cl (p.p.m.)	(9)	492	617	683	586	318	711	882	849	826	818	777	860	771	713	816	352	635	784	2,100 ± 100
x _{Cl}	(8)	792	870	933	823	474	884	1,165	1,170	1,116	1,089	1,179	1,324	1,135	1,022	1,102	551	1,037	1,124	

(1) Dome fragments: 1311E: first peléean pyroclastic flow, 1311F1 to 901F4: second peléean pyroclastic flow. V7-M: separated matrix from sample 1311F1. Density measurement systematics show a normal distribution for the first pumice fallout layer (a-d, samples 1310A1 to 1310D1) and a multi-modal distribution for the layers of the second plinian sequence composed of intercalated ash and pumice fallout and flow deposits (Fig. 2e). Several clasts from pumice fallout layer (j) have been analysed to cover the entire density range (samples 1310J1 and J5-8 to J5-70). See ref. 8 for sample description.

(2) Crystallinity (C) is expressed as the weight fraction of phenocrysts. It is estimated by mass balance calculation using groundmass and crystal separates compositions (major and trace elements) measured on two reference samples (1310C1 and 1311F). This calculation method ensures the data set consistency; it has been tested by comparing mass-balance calculations and point-counting measurements on a tenth of samples. Mean propagated 2σ error is ~5%. Groundmass is entirely glassy in plinian clasts and microcrystalline in dome fragments.

(3) Densities (d) are calculated, on a crystal-free basis, from bulk clasts densities measured by a gravimetric method. Initial melt density (d_i) is estimated by the vesicle-free melt density measured by pycnometry on finely crushed groundmasses of four reference samples and corrected for crystal content. Mean 2σ error is <5%.

(4) Vesicularity in %. It is calculated from clast and melt densities (d , d_i).

(5) Ratio of gas volume to groundmass volume. It is calculated by the formula: $V_g/V_l = (d_i/d) - 1$. Mean 2σ error is ~10%.

(6) H₂O content of the bulk samples measured by gravimetric method. Mean 2σ error is ~10%.

(7) From experimental petrology and melt-inclusion analyses¹¹.

(8) x_{H_2O} and x_{Cl} represent H₂O and Cl weight fractions in melts and are calculated by dividing bulk sample content by (1 - C) where C is crystallinity (2).

(9) Cl contents of bulk samples measured by ion selective potentiometry. Mean 2σ error is 10%. Cl content of initial melt is a mean of eight electron microprobe measurements in melt inclusions of plagioclases and pyroxenes from the last plinian layer.

Because of the very small dimensions of the glass fragments (generally <10 μ m) due to high vesicularity or microcrystallinity of the clasts, direct measurements of volatile contents of glasses are extremely difficult. The few available measurements using electron microprobe are in complete agreement with calculated x_{H_2O} and x_{Cl} values: H₂O, 0.4 ± 0.7 in glasses from dome fragments and 1.8 ± 0.3 in glasses from a plinian clast¹¹; Cl, 1,240–1,810 p.p.m. in glasses from a vesiculated clast of the last plinian layer (range of ten measurements, this work).

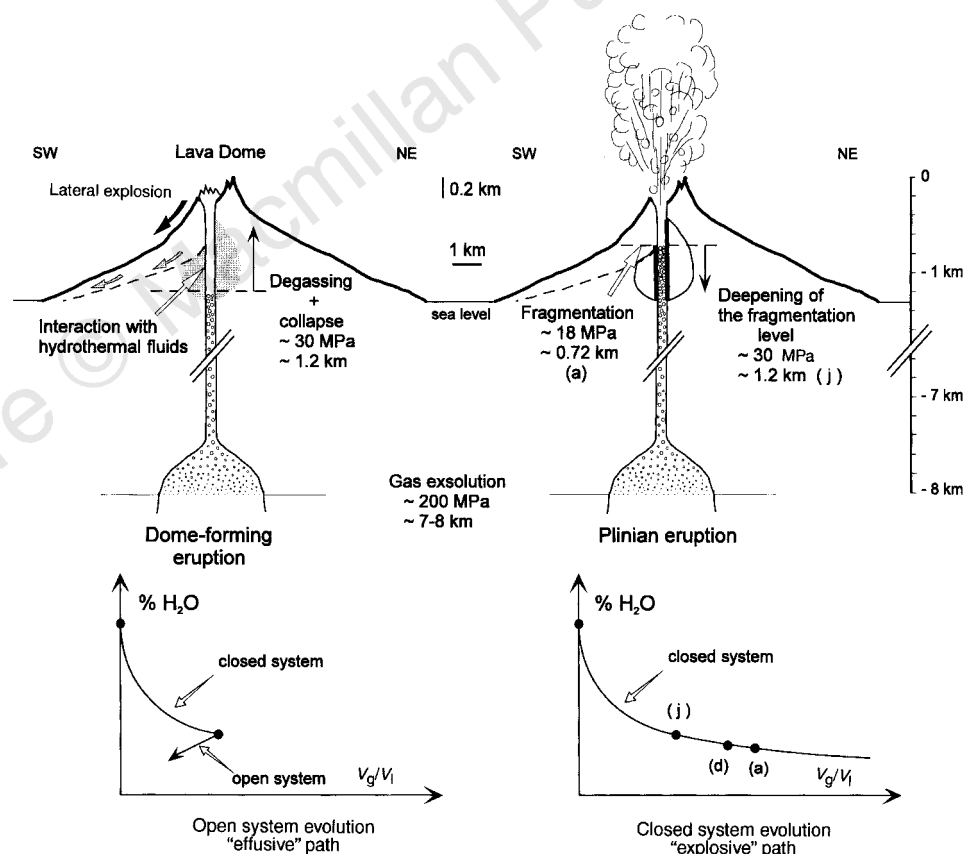


Figure 3 Sketch of the degassing evolution in a closed and in an open system for the Mt Pelée P1 eruption. On the schematic SW-NE section of Mt Pelée volcano are represented the feeding conduit and the position of the hydrothermal system. On the SW flank is represented the horse-shoe-shaped structure corresponding to an old flank collapse event (>25 kyr ago) filled by the recent pyroclastic deposits²⁵. This structure cuts the top of the hydrothermal system at ~650 m under the summit, as evidenced by active hot springs. Dashed arrows indicate the down-flow of thermal waters to the sea. The H₂O degassing paths for both eruptive styles are given in the H₂O- V_g/V_l diagrams at the bottom of the figure (see also Fig. 2). The H₂O vapour exsolution begins at ~7–8 km in the magma chamber, and increases with magma ascent in the feeding conduit. When the

magma remains in closed system (right-hand diagrams), fragmentation begins when the vesicularity is ~75% (~0.72 km) and gives rise to a plinian eruption. During the eruption the fragmentation level deepens until ~1.2 km (vesicularity ~60%). During open-system evolution (left-hand diagrams) the magma loses its gas through the permeable conduit walls and interacts with the hydrothermal system from ~1.2 km. Gas loss and crystallization lead to the modification of the magma rheology and the evolution to a dome-forming activity. The incompletely degassed inner parts of the lava dome produce lateral explosions of the dome, giving rise to peléean pyroclastic flows. The transition from open system (left) to closed system (right) can be explained by conduit sealing by the previous dome-forming activity or by an increase of the deep magma supply.

where d_{v-1}^{Cl} is the vapour–melt partition coefficient of Cl. As V_g/V_1 is only a function of x_{H_2O} , the theoretical degassing behaviour of Cl may be represented in an $x_{Cl}-V_g/V_1$ diagram (Fig. 2d).

In the $x_{H_2O}-V_g/V_1$ and $x_{Cl}-V_g/V_1$ diagrams, the clasts of the first plinian fallout layer (a–d) plot on the closed-system evolution line (Fig. 2b, d), corresponding to a fragmentation pressure of ~18–20 MPa (~0.8 km depth). The clasts of the last plinian fallout layer (j) with V_g/V_1 in the range 1.5–2.5 (~50% of the total population, Fig. 2e) also plot close to this line but at a higher pressure (~30 MPa) which suggests an increase of the fragmentation depth up to ~1.2 km (Fig. 2b–d). The most vesiculated clasts of layer (j) ($1.3 < V_g/V_1$) display a large V_g/V_1 range with a roughly constant x_{H_2O} (Fig. 2c). This could result from a continuous expansion of gas after magma fragmentation from ~30 MPa (1.2 km depth) up to ~12 MPa (0.5 km depth) for the most vesiculated clast, without any further significant H_2O degassing from the melt (Fig. 2a, c).

All dome fragments and the dense clasts from plinian layer (j) ($V_g/V_1 < 1.5$, ~25% of the clasts) plot far from the closed-system evolution line but define curved trends which intercept this line at $P < 30$ MPa. In agreement with isotopic results, this suggests that the processes of magma degassing have evolved from closed- to open-system conditions. The latter stage corresponds to the interconnection of the vesicles allowing gas to escape (x_{H_2O} decrease) from the permeable magma into surrounding rocks, according to Eichelberger's model⁴. As gas pressure in vesicles can no longer be maintained, pre-existing vesicles collapse (V_g/V_1 decrease, Fig. 2a). The ratio, K , between the decompression rate ($R_m = dP/dt$) and the vesicle collapse rate ($R_{col} = (dV_g/V_{g0})/dt$; V_{g0} is an arbitrary reference gas volume) may be estimated from the slope of the fitted evolution curve in the $x_{H_2O}-V_g/V_1$ diagram:

$$K = d(x_{H_2O})/d(V_g/V_1) = R_{degas}/(R_{col} V_{g0}/V_1) \quad (4)$$

with $R_{degas} = dx_p/dt = R_m(dx_p/dP)$.

Differentiation of equation (1) gives $dx_p/dP = x_p n_{H_2O}/P$, which finally leads to:

$$(R_m/R_{col}) = K(V_{g0}/V_1)/(x_p n_{H_2O}/P) \quad (5)$$

where R_m and R_{col} are expressed in $kbar s^{-1}$ and $vol\% s^{-1}$ respectively.

The R_m/R_{col} values for dense clasts from layer (j) (Fig. 2c) range from 7 to 0.4 for a depth varying between 1.2 km (where $V_{g0}/V_1 = 1.6$) and 0.2 km. Using two constant decompression rates of 10^{-8} and 10^{-7} $kbar s^{-1}$ (that is, magma ascent rates of 10^{-4} and 10^{-3} $m s^{-1}$; refs 2, 17, 18), the vesicle collapse rate increases from 2×10^{-5} to 2×10^{-3} and 2×10^{-4} to 2×10^{-2} $vol\% s^{-1}$ respectively. These collapse rates (0.1 to 20% per 2 h) are slower than those predicted for silicic melts in similar conditions (typically 50% per 2 h; ref. 10). However, collapse rate is inversely proportional to viscosity which, in such highly microcrystalline magmas, is much larger than that of crystal-free melts. Moreover, dramatic viscosity increases are expected¹⁹ when crystallinity exceeds ~30–40%. Considering that initial crystallinity of magmas of the P1 eruption is ~30% and that gas loss induces the nucleation of microlites²⁰, degassing under open-system conditions may drastically increase magma viscosity.

These closed (plinian type) and open (dome-forming type) system evolutions are in agreement with micro-textural observations⁹. Dome fragments are very heterogeneously vesiculated (vesicle size range, 4–300 μm) as compared to plinian clasts (~3–20 μm). Vesicle shapes are pseudo-spherical in plinian clasts while they are extremely irregular in dome fragments with evidence of flattening. The matrix is entirely glassy in plinian clasts, whereas it is highly microcrystalline in the dome fragments. All these features suggest a rapid, single-stage degassing in a closed system for plinian eruption and slow, multistep degassing path with a stage of vesicle collapse during dome-forming eruption.

We propose a likely model for this eruption (Fig. 3). From the deep magma chamber, degassing is a near-equilibrium process in

closed system. At shallow depth (~1 km), when vesicularity is large ($V_g/V_1 \approx 1.6$), vesicle interconnection occurs and gas may escape through the permeable conduit walls. This gas loss leads to nucleation of microlites²⁰ and a drastic viscosity increase which reduces the magma ascent rate and in turn, increases the gas escape efficiency. Local pressure equilibrium between magma and surroundings may thus be achieved, allowing chemical interactions between the outer zone of the magma body and the hydrothermal system, as evidenced by Th isotopes⁸. This bulk process leads to a lateral heterogeneity of the growing dome with the preservation of less-degassed inner parts and the development of sheared zones in the outer parts. This might explain the explosivity of the domes formed during the P1 eruption (typical of pelean eruptions) in response to slight gas overpressures²¹. The transition from dome-forming to plinian activity may occur only if gas loss and hydrothermal interaction are significantly reduced: this can be achieved either by an increase of the deep magma supply leading to higher ascent rates and shorter duration of the wall-rock/magma interactions, or by reduction of the permeability of the conduit walls due to sealing processes such as fracture infilling by debris or silification^{22,23}. This mechanism is very different from the various current dynamic models because it involves an active role of the surrounding fluids in determining the eruptive style. The example of P1 eruption shows that the transition between open- and closed-system evolution may occur in both directions in the same eruption. The parameters that control the eruptive style (such as magma viscosity, degassing and ascending rates, wall rock and magma permeabilities) are interdependent in a complex manner^{21,24}. A slight modification of one parameter close to some critical value may lead to an important variation of the magma rheology and then to a sharp transition in the eruptive style. This interpretation is in agreement with the occurrence of both eruptive styles in the activity at Mt Pelée (<14 kyr ago) and also with large variations of the explosivity of the growing lava domes in spite of very close magmatic compositions. □

Received 1 May 1997; accepted 5 January 1998.

- Eichelberger, J. C., Carrigan, C. R., Westrich, H. R. & Price, R. H. Non-explosive silicic volcanism. *Nature* **323**, 598–602 (1986).
- Jaupart, C. & Allègre, C. J. Gas content, eruption rate and instabilities in silicic volcanoes. *Earth Planet. Sci. Lett.* **102**, 413–429 (1991).
- Hoblitt, R. P. & Harmon, R. S. Bimodal density distribution of cryptodome dacite from the 1980 eruption of Mount St. Helens, Washington. *Bull. Volcanol.* **55**, 421–437 (1993).
- Eichelberger, J. C. Silicic volcanism: ascent of viscous magmas from crustal reservoirs. *Annu. Rev. Earth Planet. Sci.* **23**, 41–63 (1995).
- Hoblitt, R. P. et al. in *Fire and Mud Eruptions and Lahars of Mt Pinatubo, Philippines* 457–511 (Univ. Washington Press, 1996).
- Sparks, R. S. J., Gardeweg, M. C., Calder, E. S. & Matthews, S. J. Erosion by pyroclastic flows on Lascar Volcano, Chile. *Bull. Volcanol.* **58**, 557–565 (1997).
- Boudon, G., Bourdier, J. L. & Traill, H. High-energy pyroclastic flows in the recent activity of Mt. Pelée, Martinique. (abstr.) (IAVCEI, Ankara, 1994).
- Villemant, B., Boudon, G. & Komorowski, J. C. U-series disequilibrium in arc magmas induced by water-magma interaction. *Earth Planet. Sci. Lett.* **140**, 259–267 (1996).
- Villemant, B., Komorowski, J. C. & Boudon, G. Geochemical and textural insights into magma degassing processes during the Peléan-Plinian P1 eruption of Mt. Pelée. *Eos* **77**, 829 (1996).
- Gardner, J. E., Thomas, R. M. E., Jaupart, C. & Tait, S. Fragmentation of magma during Plinian volcanic eruptions. *Bull. Volcanol.* **58**, 144–162 (1996).
- Martel, C. *Conditions Pré-éruptives et Dégazage des Magmas Andésitiques de la Montagne Pelée (Martinique)* Thesis, Univ. Orléans (1996).
- Burnham, C. W. in *The Evolution of Igneous Rocks: Fiftieth Anniversary Perspectives* (ed. Yoder, H. S.) 439–482 (Princeton Univ. Press, 1979).
- Burnham, C. W. in *Volatiles in Magmas* 123–129 (Vol 30, Reviews in Mineralogy, Mineral. Soc. Am., Washington DC, 1994).
- Metric, N. & Rutherford, M. J. Experimental study of chlorine behaviour in hydrous silicic melts. *Geochim. Cosmochim. Acta* **56**, 607–616 (1992).
- Webster, J. D. Water solubility and chlorine partitioning in Cl-rich granitic systems: effects of melt composition at 2 Kbar and 800 °C. *Geochim. Cosmochim. Acta* **56**, 679–687 (1992).
- Barclay, J., Carroll, M. R., Houghton, B. F. & Wilson, C. N. J. Pre-eruptive volatile content and degassing history of an evolving peralkaline volcano. *J. Volcanol. Geotherm. Res.* **74**, 75–87 (1996).
- Lacroix, A. *La Montagne Pelée et ses Eruptions* (Masson, Paris, 1904).
- Sato, H., Fujii, T. & Nakada, S. Crumbling of dacite dome lava and generation of pyroclastic flows at Unzen volcano. *Nature* **360**, 664–666 (1992).
- Lejeune, A.-M. & Richet, P. Rheology of crystal bearing silicate melts: an experimental study at high viscosities. *J. Geophys. Res.* **100**, 4215–4229 (1995).
- Swanson, S. E., Naney, M. T., Westrich, H. R. & Eichelberger, J. C. Crystallization history of Obsidian dome, Inyo Domes, California. *Bull. Volcanol.* **51**, 161–176 (1989).
- Sparks, R. S. J. Causes and consequences of pressurisation in lava dome eruptions. *Earth. Planet. Sci. Lett.* **150**, 177–189 (1997).
- Heiken, G., Wohletz, K. & Eichelberger, J. Fracture fillings and intrusive pyroclasts, Inyo Domes, California. *J. Geophys. Res.* **93**, 4335–4350 (1988).

23. Komorowski, J. C. *Scanning Electron Microscopy of Pyroclastic Matter: Eruptions of Mt Vesuvius in AD 79 and Mt St Helens in AD 1980*. Thesis, Arizona State Univ. (1991).
24. Woods, A. W. & Koyaguchi, T. Transitions between explosive and effusive eruptions of silicic magmas. *Nature* **370**, 641–644 (1994).
25. Vincent, P. M., Bourdier, J.-L. & Boudon, G. The primitive volcano of Mt Pelée: its construction and partial destruction by flank collapse. *J. Volcanol. Geotherm. Res.* **38**, 1–16 (1989).

Acknowledgements. We thank J. C. Komorowski and R. Clocchiatti for assistance with analyses; P. Agrinier for communicating δD analytical results; S. Sparks for reviews; and C. Jaupart, N. Metrich, J. C. Komorowski, M. P. Semet and S. Tait for discussions. This work was supported by the Programme National sur les Risques Naturels (CNRS-INSU).

Correspondence and requests for materials should be addressed to B.V. (e-mail: villemant@ipgp.jussieu.fr).

Selectivity of extinction among sea urchins at the end of the Cretaceous period

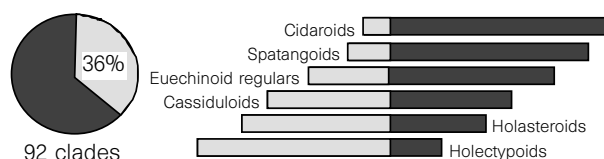
Andrew B. Smith & Charlotte H. Jeffery

Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

By compiling large databases and searching for environmental and palaeobiological correlates associated with survival, insight can be gained into the driving mechanisms involved in mass extinctions^{1–4}. Although this approach lacks precise temporal resolution and thus cannot be used to investigate how rapidly extinction took place, it provides a broad overview, less plagued by sampling problems caused by shifting facies. Here we present a global analysis of a major marine invertebrate group, the sea urchins, which suffered 36% extinction at genus level in the late Maastrichtian age and continuing high levels of extinction in the Danian age. No preferential survivorship was found for clades with widespread distribution, but there was a strong correlation between feeding strategy and survivorship at the end of the Cretaceous period. Surprisingly, however, clades whose larvae must feed to reach metamorphosis were not significantly harder hit than those with non-feeding larval development. Our results indicate that nutrient supply was a crucial factor in driving K/T-boundary extinctions, with selection more strongly focused on benthic adult than on larval planktotrophic stages.

We have revised and standardized the taxonomy of Maastrichtian and Palaeocene echinoids worldwide to generate a cladistic phylogeny for the group at species level⁵. We recognize ~250 Maastrichtian echinoid species, placed into 115 approximately genus-level clades. Of 92 clades that are definitely recorded from late Maastrichtian

Upper Maastrichtian/Danian



Danian/Thanetian

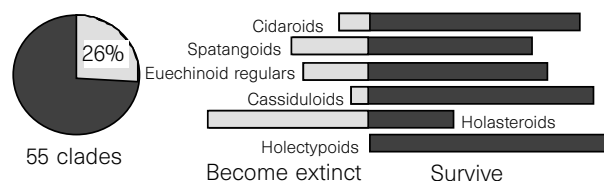


Figure 1 Genera of echinoids with last occurrences in the late Maastrichtian and Danian. Extinctions as proportions of total diversity (pie charts) and relative extinction levels for major higher taxa (bar charts) are shown.

Box 1 Correlates of survivorship for echinoids at the end of the Cretaceous period

Statistically significant correlation links survival with:

- Taxonomic structure (Fig. 2a)
- Feeding strategy (Fig. 2b)
- Palaeogeographic region (endemics only), with extinction levels highest in the Americas and lowest in the Indo-Pacific (Fig. 2c)

Survivorship weakly correlated with:

- Absolute palaeolatitude (in degrees away from the equator), with higher extinction levels encountered at low palaeolatitudes (Fig. 2e)
- Absolute palaeolongitude, with extinction levels lowest to the east (Fig. 2o)
- Water depth (Fig. 2d), with shallow-water faunas harder hit than deeper-water faunas. Analysis suggests that extinction in shallow water is concentrated in carbonate-platform facies
- Larval strategy: whether the organism is planktotrophic (and obliged to feed in the plankton before metamorphosis) or non-planktotrophic (Fig. 2g)

Survivorship independent of:

- Numbers of Maastrichtian morphospecies included in genus (Fig. 2j)
- Geographical distribution before the K/T event, measured by palaeogeographic range (Fig. 2i, l) and the numbers of discrete localities (Fig. 2k) or large-scale provinces (Fig. 2m) in which a clade is recorded
- Whether species were epifaunal or infaunal (Fig. 2f)
- Whether food was collected by grazing/scavenging or by deposit feeding (Fig. 2h)

strata, 36% have no post-Maastrichtian record. Further significant extinction occurred during the Danian (early Palaeocene), however, with 26% of the 55 clades present making their last appearance (Fig. 1). We collected environmental data on substrate, inferred water depth and palaeogeographic distribution, and palaeobiological data on maximum size, mode of life, feeding strategy and diversity for each occurrence. Evidence for selectivity of extinction was sought by testing for correlation between survivorship and these biological and geographical factors (Fig. 2). Box 1 summarizes our findings.

Survival was independent of species diversity within each clade (Fig. 2j) and thus is not a taxonomic artefact determined by the size of the taxonomic units termed 'genus'. It was also largely independent of the geographical range occupied during the Maastrichtian (Fig. 2i, l and m). It has previously been suggested that widespread genera preferentially survive at mass-extinction events^{1,6}. Our data differ, however, probably because we have constructed an explicitly phylogenetic framework that identifies clade survival independently of taxonomic names. In our raw data we found many wrongly identified taxa recorded from single formations, inflating the estimate of 'endemic' extinctions.

Slight preferential losses of low-latitude faunas (Fig. 2e) and of shallow-water faunas (Fig. 2d) are inextricably linked, as deep-water sediments containing echinoid faunas from low latitudes are rare. Low-latitude extinctions are largely explained by the virtual disappearance of shallow carbonate-platform facies in the Danian, as is the difference in survivorship between Western Tethyan and Eurasian faunas (Fig. 2c). Loss from shallow-water carbonate settings took place across all habitats (Fig. 2n) and is not simply restricted to perireefal settings.

There is statistical support for extinction having been most intense in the Americas and lowest in the Indo-Pacific (Fig. 2c, o) when considering endemic species, but our numbers are extremely low and there are two confounding problems: a dearth of Danian echinoid faunas from the Americas, and uncertainty over the precise

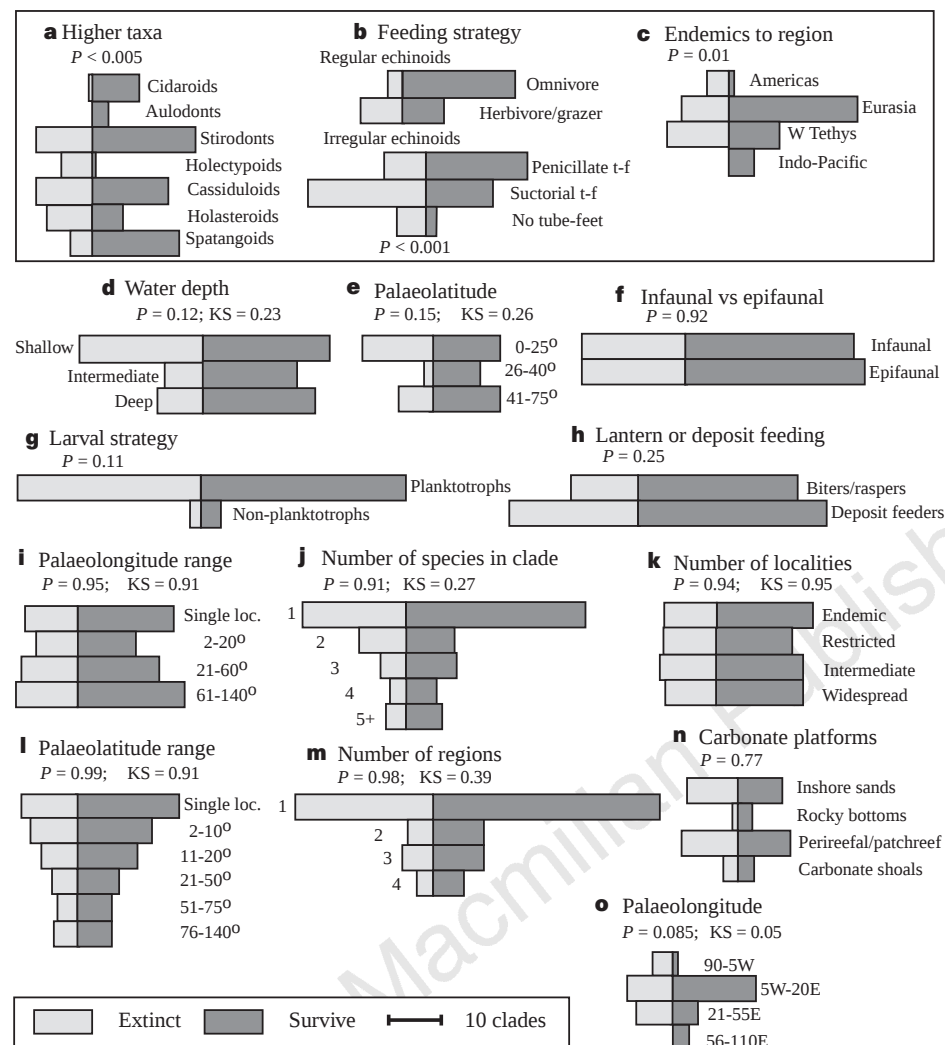


Figure 2 Proportions of Maastrichtian clades surviving into the Danian or becoming extinct. They have been partitioned according to: **a**, higher taxonomic grouping; **b**, feeding strategy, as reflected by the structure and arrangement of tube-feet (t-f) around the mouth; **c**, taxa restricted to a single geographic region; **d**, inferred water depth occupied ('shallow' corresponds to within fair-weather wave base; 'intermediate' corresponds to within storm wave base; 'deep' corresponds to basinal muds and chalks); **e**, palaeolatitude (taxa restricted to one band only) (in degrees away from palaeo-equator); **f**, infaunal or epifaunal habit; **g**, planktotrophic or non-planktotrophic larval stage; **h**, grazing or deposit feeding; **i**, palaeolongitudinal range (in degrees) ('single loc.' represents taxa recorded from one locality only); **j**, number of species in each clade; **k**, number of localities the clade is reported from (an endemic clade is present in 1 locality, a restricted clade in 2–4 localities, an intermediate clade in 5–8 localities and a widespread clade in more than 8 localities); **l**, palaeolatitudinal range (in degrees); **m**, number of regions (Americas, Eurasia, Western Tethys, Indo-Pacific) the clade occurs in; **n**, habitat for carbonate platform taxa; **o**, palaeolongitude (taxa restricted to one band only). P represents significance level found with a chi-squared test; KS represents the significance level found in a Kolmogorov–Smirnov test.

age of Cuban late Cretaceous endemic species (here treated as Maastrichtian). Further data are needed to decide whether this pattern is genuine or an artefact of sampling.

Where facies continuity exists across the K/T boundary, as in Eurasian chalk facies, there was considerably less extinction. Although it has been suggested^{6,7} that chalk faunas were severely affected at the end of the Cretaceous, our evidence suggests that the main bout of extinction for chalk echinoids took place towards the end of the Danian when chalk deposition ended.

Levels of extinction at or near the end of the Cretaceous vary significantly amongst higher taxa (Fig. 2a), suggesting that differences in palaeobiology may have been important. We found that feeding strategy was highly correlated with survival (Fig. 2b). All irregular echinoids are deposit feeders, collecting detritus from the sea floor either by directly funnelling sediment into the peristome without the aid of tube feet or by selectively picking up particles from the sea floor using perioral tube feet⁸. Holasteroids and spatangoids have highly specialized penicillate tube feet that allow them to manipulate fine detrital material (of $<200 \mu\text{m}$ in size) (refs 8, 9). Cassiduloids and holotypoids have only simple, suckered tube feet and are restricted to feeding on larger particles (of ~ 500 – $1,500 \mu\text{m}$ in size) (refs 10, 11). Echinoids lacking tube feet continuously ingest the surface layer of sediment in nutrient-poor settings, passing large quantities through their gut to obtain adequate nutrition. Deposit feeders possessing penicillate tube feet and able to select fine organodetritus show significantly enhanced survivorship compared with the other types of feeders

(Fig. 2b), with extinctions concentrated in nutrient-starved inner-shelf carbonate-platform settings.

It is more difficult to recognize discrete feeding strategies among regular echinoids, all of which feed by means of their lantern. Most are opportunistic, capable of feeding on various substrates⁸. Those with distinct phyllodes and keeled teeth are specialist rasps and grazing herbivores, however, and suffered significantly higher levels of extinction at the end of the Cretaceous than generalist omnivores without phyllodes. Deposit feeders as a whole were hardly more affected than lantern feeders (Fig. 2h).

Surprisingly, although clades with planktotrophic larvae appear somewhat more prone to extinction than clades with non-feeding (non-planktotrophic) larvae (Fig. 2g), the difference is not statistically significant, as was also found for gastropods¹². As the majority of echinoids and gastropods with obligate planktotrophic larvae survived into the Tertiary period, it seems unlikely that the differential fates of ammonites and nautiloids at the K/T boundary can be ascribed directly to a difference in their larval development, as has previously been suggested⁶. These results also discredit bolide scenarios that invoke an instantaneous catastrophe involving the wholesale selective extermination of planktotrophs.

The correlation between feeding strategy and survival at the end of the Cretaceous is *prima facie* evidence that extinction of echinoids was in some way nutrient-driven, but why should such extinction be focused more strongly on benthic adults than on their planktonic larvae? The food supply for deposit feeders fluctuates because of seasonality of the phytoplankton, leading to

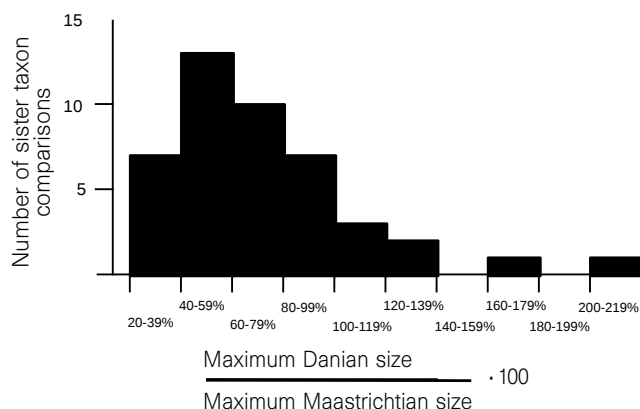


Figure 3 Difference in maximum organism sizes between Maastrichtian and Danian sister taxa.

nutrient stress throughout much of the year^{13,14}. Furthermore, the proportion of phytodetritus reaching the sea floor rapidly decreases below the euphotic zone¹⁴. The observed pattern could, therefore, have been produced by a decrease in phytoplankton abundance at the end of the Cretaceous that was not so large as seriously to affect planktotrophic larvae living and feeding in the euphotic zone, but which reduced the organic matter reaching the sea floor sufficiently to trigger widespread extinction of already nutrient-stressed deposit feeders.

It is possible that the final blow was dealt by asteroid impact, but there is indirect evidence that conditions for plankton were becoming less favourable immediately before the K/T boundary. Climate was rapidly deteriorating¹⁵ and extinction of several major molluscan groups had already taken place¹⁶. Furthermore, numerous lineages of echinoids independently switched to non-planktotrophic development in the Maastrichtian, regardless of palaeolatitude and water depth¹⁷, implying that survival for planktonic feeding larvae was becoming markedly less predictable. Furthermore, the fact that high levels of extinction continued into the Danian suggests a slow squeeze rather than an instantaneous catastrophe.

Finally, we note a dramatic decrease in size of post-Cretaceous survivors. Almost all Danian echinoids are significantly smaller than their Maastrichtian antecedents (Fig. 3) and apparently remained so until the latter half of the Danian. Early Danian echinoids either grew much more slowly or became more opportunistic, achieving sexual maturity at a much earlier stage. In either case, the small size of Danian survivors is consistent with nutrient supply remaining unpredictable and a limiting factor to growth for a considerable time interval following the K/T event, as postulated previously¹⁸. □

Received 9 July; accepted 17 November 1997.

- Jablonski, D. Mass extinctions: persistent problems and new directions. *Geol. Soc. Am. Spec. Pap.* **307**, 1–9 (1996).
- Jablonski, D. in *Dynamics of Extinction* (ed. Elliott, D. K.) 183–229 (Wiley, New York, 1986).
- Jablonski, D. & Raup, D. M. Selectivity of end-Cretaceous marine bivalve extinction. *Science* **268**, 389–391 (1995).
- Jablonski, D. Progress at the K–T boundary. *Nature* **387**, 354–355 (1997).
- Smith, A. B. & Jeffery, C. H. Fossil echinoids of the Maastrichtian and Palaeocene: an illustrated key. *Spec. Pap. Palaeont.* (submitted).
- Sheehan, P. M., Coorrough, P. J. & Fastovsky, D. E. Biotic selectivity during the K/T and late Ordovician extinction events. *Geol. Soc. Am. Spec. Pap.* **307**, 477–489 (1996).
- Johansen, M. B. Background extinction and mass extinction of the brachiopods from the chalk of northwest Europe. *Palaios* **4**, 243–250 (1989).
- De Ridder, C. & Lawrence, J. M. in *Echinoderm Nutrition* (eds Jangoux, M. & Lawrence, J. M.) (Balkema, Rotterdam, 1982).
- Smith, A. B. The structure, function and evolution of tube feet and ambulacral pores in irregular echinoids. *Palaeontology* **23**, 39–83 (1980).
- Telford, M. & Mooi, R. Podial particle picking in *Cassidulus caribaeorum* (Echinodermata: Echinoidea) and the phylogeny of sea urchin feeding mechanism. *Biol. Bull.* **191**, 209–223 (1996).
- Rose, E. P. F. Some observations on the recent holcypoid echinoid *Echinoneus cyclostomus* and their palaeoecological significance. *Thalass. Jugoslav.* **12**, 299–306 (1978).
- Valentine, J. W. & Jablonski, D. Mass extinctions: sensitivity of marine larval types. *Proc. Natl Acad. Sci. USA* **83**, 6912–6914 (1986).

- Levinson, J. S. Trophic group and the end-Cretaceous extinction: did deposit feeders have it made in the shade? *Paleobiology* **22**, 104–112 (1996).
- Summerhayes, C. P. & Thorpe, S. A. *Oceanography, an Illustrated Guide* (Manson, London, 1996).
- Barrera, E. Global environmental change preceding the Cretaceous–Tertiary boundary: Early–late Maastrichtian transition. *Geology* **22**, 877–880 (1994).
- Macleod, N. et al. The Cretaceous–Tertiary biotic transition. *J. Geol. Soc. Lond.* **154**, 265–292 (1997).
- Jeffery, C. H. Dawn of echinoid nonplanktotrophy: coordinated shifts in development indicate environmental instability prior to the K–T boundary. *Geology* **25**, 991–994 (1997).
- D'Hondt, S., Herbert, T. D., King, J. & Gibson, C. Planktonic foraminifera, asteroids, and marine production: death and recovery at the Cretaceous–Tertiary boundary. *Geol. Soc. Am. Spec. Pap.* **307**, 303–317 (1996).

Acknowledgements. We thank M. Foote, J. Young, C. Marshall, P. Ward and J. Levinson for their comments on this paper. Research was supported by a NERC research grant.

Correspondence and requests for materials should be addressed to A.B.S. (e-mail: a.smith@nhm.ac.uk).

A molecular evolutionary framework for the phylum Nematoda

Mark L. Blaxter*, Paul De Ley†‡, James R. Garey§, Leo X. Liu¶, Patsy Scheldeman†, Andy Vierstraete†, Jacques R. Vanfleteren†, Laura Y. Mackey#, Mark Dorris*, Linda M. Frisse*, J. T. Vida* & W. Kelley Thomas*

* Institute of Cell, Animal and Population Biology, King's Buildings, University of Edinburgh, Edinburgh EH9 3JT, UK

† Department of Morphology, Systematics and Ecology, University of Gent, B-9000 Gent, Belgium

‡ International Institute of Parasitology, St Albans AL4 0XU, UK

§ Department of Biological Sciences, University of South Florida, Tampa, Florida 33620, USA

¶ Harvard–Thorndike Laboratory, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA

¶ NemaPharm, Inc, 26 Landsdowne Street, Cambridge, Massachusetts 02139, USA

Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, Texas 77030, USA

* Division of Molecular Biology and Biochemistry, School of Biological Sciences, University of Missouri, Kansas City, Missouri 64110, USA

Nematodes are important: parasitic nematodes threaten the health of plants, animals and humans on a global scale^{1,2}; interstitial nematodes pervade sediment and soil ecosystems in overwhelming numbers³; and *Caenorhabditis elegans* is a favourite experimental model system⁴. A lack of clearly homologous characters and the absence of an informative fossil record have prevented us from deriving a consistent evolutionary framework for the phylum. Here we present a phylogenetic analysis, using 53 small subunit ribosomal DNA sequences from a wide range of nematodes. With this analysis, we can compare animal-parasitic, plant-parasitic and free-living taxa using a common measurement. Our results indicate that convergent morphological evolution may be extensive and that present higher-level classification of the Nematoda will need revision. We identify five major clades within the phylum, all of which include parasitic species. We suggest that animal parasitism arose independently at least four times, and plant parasitism three times. We clarify the relationship of *C. elegans* to major parasitic groups; this will allow more effective exploitation of our genetic and biological knowledge of this model species.

To study the evolutionary relationships within the phylum, we constructed a database of small subunit (SSU) sequences from 53 taxa, including 41 new sequences^{5–9}. Species were chosen to cover all the major parasitic and free-living taxonomic groups. Sequences were aligned with reference to a secondary-structure model⁵ and on the basis of similarity⁸. Model phylogenies were evaluated under the criteria of maximum parsimony (MP), maximum likelihood (ML)

and minimum evolution using the neighbour-joining (NJ) method¹⁰, and were tested for statistical support with careful consideration of unequal rate effects¹¹. The consensus of the shortest MP trees found is shown in Fig. 1. The overall consensus of our analyses is shown in Fig. 2a and can be compared with a representation of previous hypotheses, shown in Fig. 2b. Our hypothesis is based on a study of a single genetic locus but permits, for the first time, the comparison of all taxa using the same defined measurement.

Nematoda is traditionally divided into two classes, namely, the predominantly terrestrial Secernentea and the predominantly marine Adenophorea. Contrary to this view, our analyses indicate

that the Adenophorea may be paraphyletic, as it includes the ancestors of the Secernentea. Phylogenies in which the Adenophorea is monophyletic by outgroup rooting score significantly worse in all methods of analysis. This result is supported by some morphological analyses which indicate the lack of synapomorphies (common characters) for Adenophorea¹² or place Secernentea as a monophyletic clade derived from adenophorean ancestry¹³. Two purely adenophorean clades are strongly supported (clades I and II in Fig. 2). Clade I groups the vertebrate-parasitic order Trichocephalida with the insect-parasitic Mermithida, plant-parasitic Dorylaimida and free-living Mononchida. Clade II links the plant-parasitic Triplonchida with the free-living Enoplida. Our analyses do not place the orders

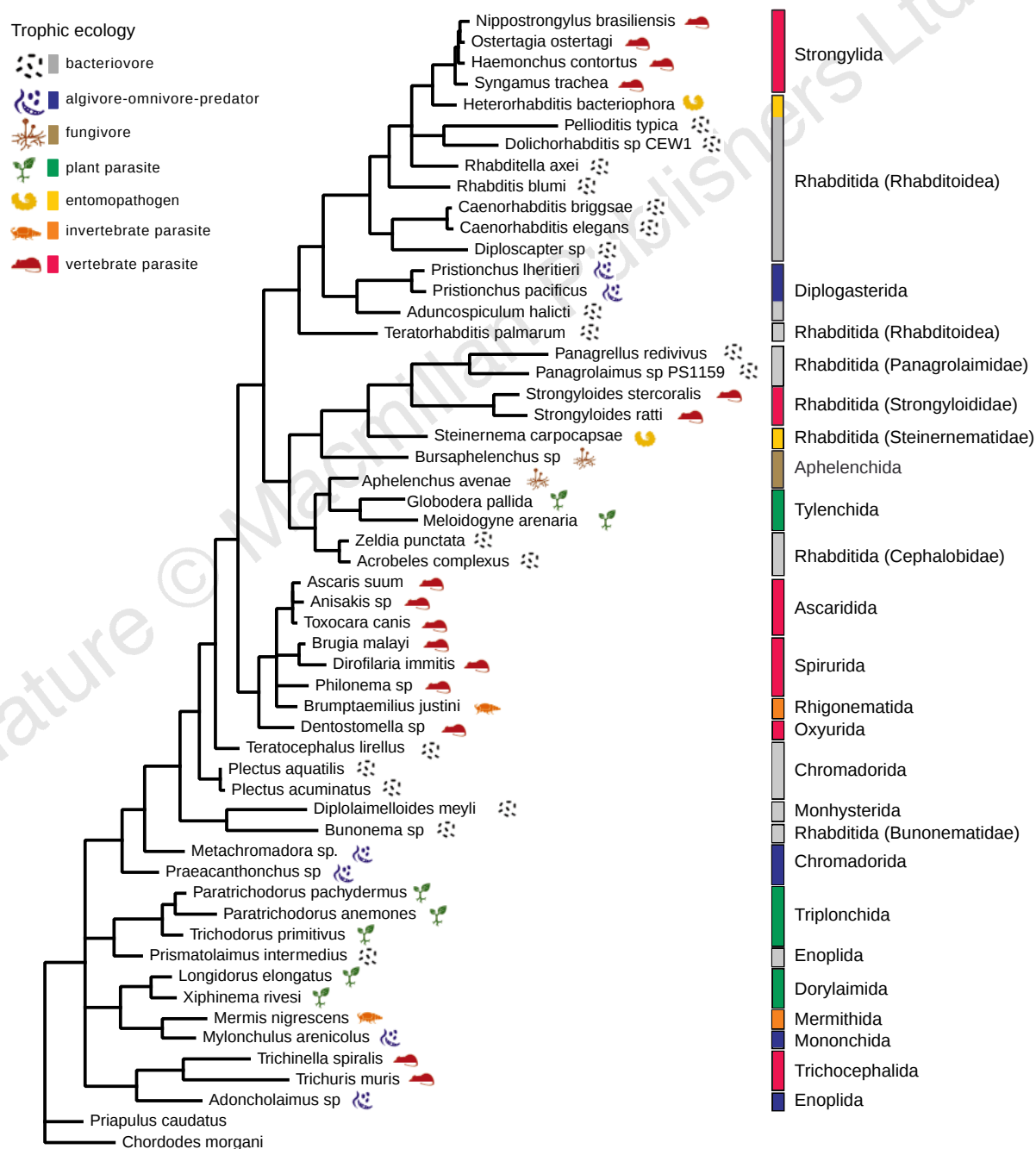


Figure 1 MP analysis of SSU sequences from 53 nematode taxa. Nearly complete SSU sequences (see Methods) were aligned and processed for MP analysis. Twenty-four shortest trees (requiring 3,583 nucleotide changes) were found; the tree presented is the strict consensus of these. Branch lengths are drawn to be proportional to the number of changes inferred. The trophic ecologies of the taxa

are represented by coloured icons. No shorter trees were revealed by local searching of subsets of the data²⁴. *Priapulius* (a priapulid) and *Chordodes* (a nematomorph) were defined as outgroups^{11,21}. In this tree, as with all MP trees constructed with subsets of the data, *Plectus* is a sister taxon to the Secernentea (see Fig. 2).

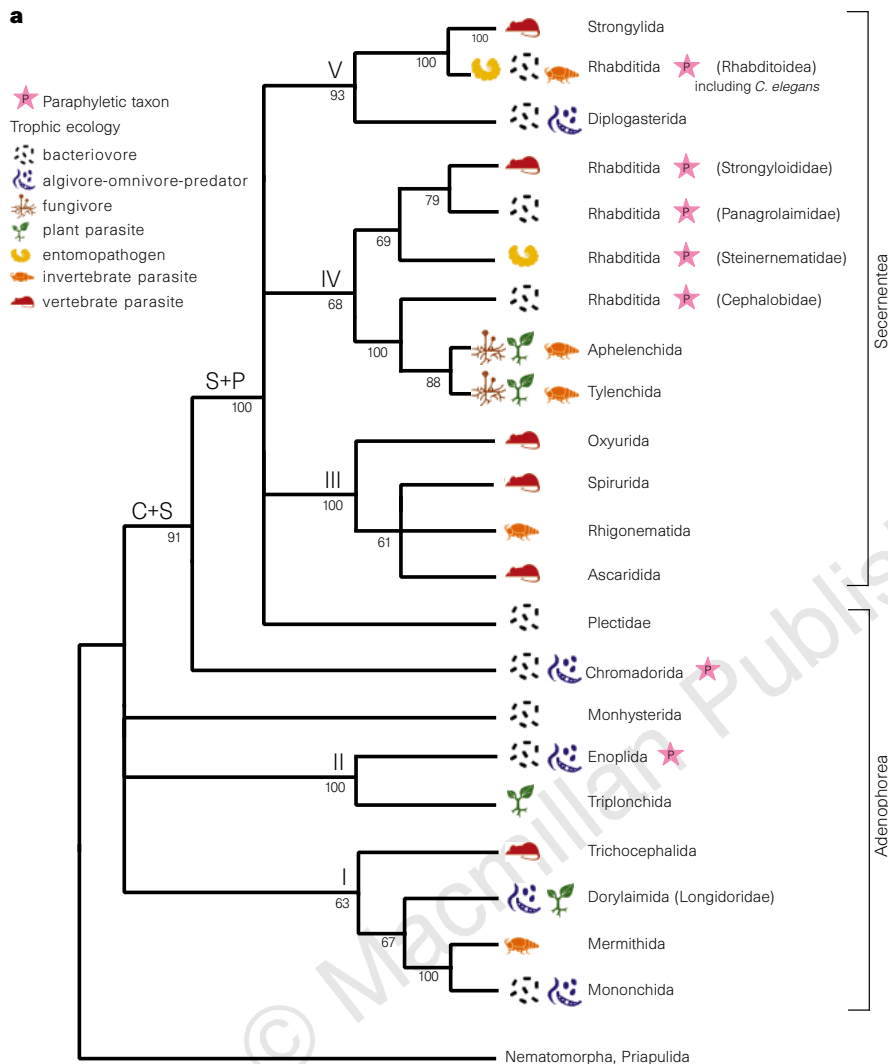
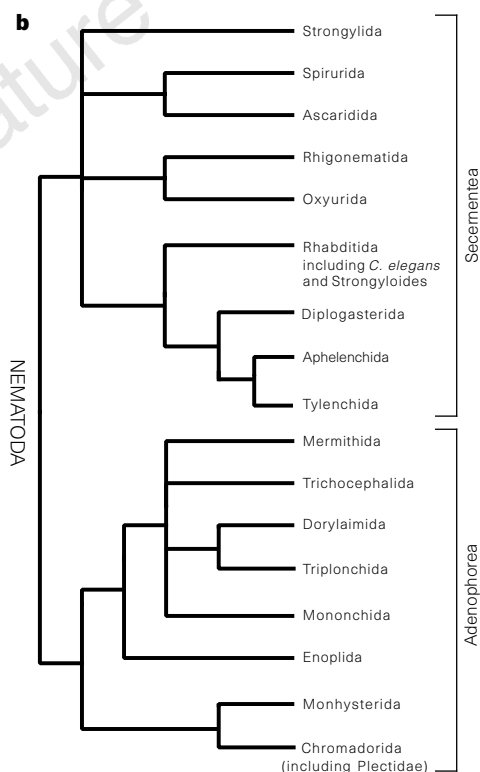


Figure 2 A phylogenetic hypothesis for the phylum Nematoda based on the SSU-sequence dataset. **a**, A dendrogram summarizing the results of our MP and NJ analyses. For each branchpoint, the maximal percentage bootstrap support in NJ analyses is given. We define five major clades in the Nematoda (from bottom, I–V). C + S indicates the clade Chromadorida plus Secernentea, and S + P indicates Secernentea plus Plectidae. Trophic ecologies are indicated as in Fig. 1. Currently accepted taxa that we find to be paraphyletic are indicated with a starred P. **b**, Our interpretation of the currently accepted systematics of the orders and genera considered in this study, represented as a cladogram and derived from several sources^{2,12–14}.



Triplonchida and Dorylaimida as sister taxa, in contrast to the results of previous studies^{12–14}. The single sampled representative of the order Monhysterida does not give an unequivocal position for this group.

The Chromadorida are a large and diverse group of mostly microbivorous nematodes, which have been proposed to include the ancestors of the terrestrial Secernentea¹³. The SSU data do not resolve the chromadorids as a distinct clade, but show a paraphyletic, comblike series of taxa at the base of the Secernentea. One chromadorid group, the Plectidae, is significantly associated with the secernentean clades (clade S + P of Fig. 2), suggesting that it may be a surviving representative of the group from within which the secernentean radiation began.

Most taxonomies divide the Secernentea into nine orders, with five of these including exclusively animal-parasitic taxa. We identify three major clades within the Secernentea (clades III, IV and V of Fig. 2); these clades do not correspond to the divisions of classical taxonomy. Instead, plant- and animal-parasitic orders are integrated within a radiation of free-living taxa. Clade III represents a novel grouping of vertebrate- and arthropod-parasitic taxa from the orders Ascaridida (large gut roundworms), Spirurida (filarial nematodes), Oxyurida (pinworms) and Rhigonematida (millipede-gut parasites). The order Rhabditida, a trophically diverse assemblage including some entomopathogenic and parasitic members, is paraphyletic, as 'rhabditid' taxa are found in two strongly supported clades, each of which includes a major parasitic order.

Clade IV (Fig. 2), a 'cephalobid' clade, groups the plant-parasitic orders Tylenchida (represented here by the cyst nematode genus *Globodera* and the root-knot nematode genus *Meloidogyne*) and Aphelenchida, the vertebrate-parasitic genus *Strongyloides* and the entomopathogenic genus *Steinernema* with free-living bacteriovores of the rhabditid families Cephalobidae and Panagrolaimidae. The second clade (clade V of Fig. 2) groups *C. elegans* and other members of the suborder Rhabditina with the vertebrate-parasitic order Strongylida (small gut nematodes, including hookworms), the entomopathogenic genus *Heterorhabditis* and the order Diplogasterida. Diplogasterida, which includes *Pristionchus pacificus*, appears to be a sister group of some rhabditids, an observation supported by other molecular evidence¹⁵. *P. pacificus* has been proposed as a comparative model for *C. elegans*¹⁶ and the relationship between these two taxa is now clear.

Although nematode parasites are generally assumed to have evolved from free-living ancestors, the precise origin and free-living sister taxa of each parasitic group are unknown. It has been suggested that all nematode parasites of animals are sufficiently closely related to be placed in the same subclass¹². From our analysis we predict multiple origins for animal-parasitic taxa throughout the phylum. A minimum of four independent origins of vertebrate parasites is supported by bootstrap analysis (Fig. 2). In keeping with the view that vertebrate parasites evolved from arthropod-parasitic ancestors², each of the secernentean vertebrate-parasitic clades is associated with arthropod-parasitic or -pathogenic taxa (*Heterorhabditis* is associated with Strongylida; *Steinernema* is associated with Strongyloides; and Rhigonematida is associated with Ascaridida, Spirurida and Oxyurida). As there are extra vertebrate- and invertebrate-parasitic groups that have not yet been sampled in our analyses², the estimate of four independent origins is a minimum. Similarly, plant parasitism seems to have evolved independently three times in Nematoda, twice in the Adenophorea (the Dorylaimida and the Triplonchida) and once in the Secernentea (the Tylenchida).

Inconsistencies between this SSU phylogeny and earlier morphological ones may partly arise because of convergent evolution. The genera *Steinernema* and *Heterorhabditis* share the unusual strategy of entomopathogenesis (after infecting insect hosts, the nematodes release toxic bacterial symbionts which quickly kill the host), but they do not share an exclusive common ancestry¹⁷. Similarly, some plant parasites in the orders Dorylaimida and Triplonchida transmit important plant viruses, yet these two orders are not closely related¹⁸. The pharynx of members of the orders Tylenchida and Aphelenchida resembles that of members of the free-living order Diplogasterida, suggesting a close relationship¹⁹. Our analysis places these two groups in different clades, suggesting two independent origins of this pharyngeal type.

Parasitic taxa are typically difficult to culture and analyse independently of their hosts. On the basis of our SSU data we can suggest free-living sister taxa that may act as good models for parasitic groups. Although *Strongyloides* is often thought to be related to rhabditids such as *C. elegans*², *Panagrellus* and its relatives may actually provide better comparative models for *Strongyloides* and *Steinernema*. *C. elegans* is, in fact, much more closely related to *Heterorhabditis* and vertebrate parasites of the order Strongylida than to any other parasitic clade, and should be a better model for these taxa. Similarly, the family Cephalobidae, which includes *Acrobeles* and *Zeldia*, could provide good models for the plant-parasitic Tylenchida. These affinities between free-living and parasitic taxa are not apparent from existing classifications, although they do agree well with some morphological features (for example, the cellular structure of the posterior stoma region²⁰ and the occurrence of a male bursa with rays in Rhabditina–Strongylida, clade V). Our trees do not unequivocally indicate a closest free-living sister taxon for the compound animal-parasitic clade II. Plectididae is their sister taxon in some analyses, contrasting with

the placement of *Plectus* basal to the Secernentea on the basis of multiple morphological characters¹².

It has been suggested that the SSU rDNA of *C. elegans* has evolved at a faster rate than usual in Metazoa²¹. Our dataset includes taxa with both fast (long branch length) and slow evolution rates. SSU sequences of the morphologically similar, free-living Rhabditida are highly divergent but the morphologically disparate, vertebrate-parasitic Strongylida show very little divergence. This indicates either that these groups have been classified previously at conflicting taxonomic levels, or that the parasitic groups have radiated relatively recently. The extent of sequence divergence between the vertebrate-parasitic taxa within clades III and V is of the same order as that of their hosts, but it is probably premature to apply host phylogenetic timing to the nematodes. Molecular phylogenetic inference can be used to provide an independent evolutionary framework with which to interpret nematode evolution. One of the most interesting implications of the SSU phylogeny is that parasitism has arisen many times from different free-living groups. By suggesting a relevant set of marker taxa, this evolutionary framework should lead to a structured approach to enable the exploitation of our knowledge of *C. elegans*. It will directly aid the investigation of the dynamics of morphological and developmental evolution^{8,16} and the biology of parasitism. The generation of genome datasets from additional nematode species²² will further aid this process. □

Methods

Isolation and sequencing of SSU genes. SSU genes were amplified by the polymerase chain reaction (PCR) from individual identified nematodes or from nematode DNA samples using universal SSU primers located close to the 5' and 3' ends. When possible, we determined the SSU sequence directly from the PCR product using a set of internal conserved primers which give complete coverage of both strands of the amplified fragment. When this approach did not yield satisfactory sequences, we sequenced cloned products from both strands. Several sequences were determined following an alternate protocol⁸. The sequences determined for this study are available in GenBank under the accession numbers U81506, U81574, U81575, U81579, U81581, U81582, AFO36586–AFO36612 and AFO36637–AFO36644. Details of material sources are available at http://www.ed.ac.uk/~mbx/nematode_ssu.html.

Alignment and phylogenetic analysis. We aligned the SSU sequences using a secondary-structure model⁵. The alignment contains 1,146 characters (excluding gaps), of which 549 are informative for MP. Alternative alignments, including one in which any region involving an insertion or deletion was removed, gave nearly identical results, and did not change the deeper branching orders. The alignment was subjected to: MP analysis using PAUP v3.1.1 (ref. 23) and MacClade²⁴; ML analysis of subsets and NJ analysis using gamma-corrected Kimura distances with MEGA v1.0 (ref. 25); and NJ analysis with gamma-corrected Kimura distances and substitution-rate compensation using Treecon v1.2 (refs 26, 27). Alpha values for gamma correction were estimated with PAML v1.2 (ref. 28) and Treecon v1.2; tree topology did not change significantly for alpha values between 0.3 and 1.0. To account for variable nucleotide composition between taxa, we performed log-determinant transformation of pairwise distances using test version 4.0d59 of PAUP*, written by D. L. Swofford. Constant sites were eliminated to minimize among-site rate heterogeneity. NJ analysis of log-determinant-transformed data yielded a tree that was not significantly different from those produced by other NJ or MP analyses. We assessed statistical support using the bootstrap-resampling technique (200 bootstrap resamplings).

Rooting the tree. We used SSU sequences from a priapulid worm and a nematode as outgroup taxa; these represent the most closely related pseudocoelomate phyla¹¹. A number of outgroup sequences from other pseudocoelomate phyla²¹ was used with no effect on the internal structure of the trees. The placement of the root was supported in both MP and NJ trees. The only consistent difference between the two methods of analysis was in the placement of *Plectus*. In NJ analyses *Plectus* is the sister taxon to clade III of Fig. 2, with strong (>95%) bootstrap support. In MP analyses *Plectus* is placed basal to all Secernentea, but this position is not supported above 50% in

bootstrap replicates. ML analyses of these two trees showed that they are not significantly different.

Effects of long-branch taxa. To identify taxa with long apparent branch lengths, we performed a four-taxon NJ analysis (using gamma-corrected Kimura distances) using *Tripedalia* (a diploblast), *Antedon* (a deuterostome) and *Glycera* (a protostome) with each nematode taxon in turn. We recorded the inferred distance from the protostome–nematode node to the nematode taxon. MP distances were derived from the phylogeny presented in Fig. 1. These long-branch-length taxa often have extreme base-composition biases, but not all taxa with extreme base compositions have long branch lengths (for example, *Brugia* has an AT content of 79% but one of the shorter inferred branch lengths). NJ and MP analyses were re-performed with successive trimming of the long-branch taxa (distance >0.19 from root in four-taxon NJ analysis) from the dataset. As would be expected²⁹, exclusion of these taxa had effects on the bootstrap support for some clades. In particular, re-analysis excluding one or all of *Panagrellus*, *Panagrolaimus*, and *Strongyloides* yielded stronger bootstrap support for the cephalobid–steinernematid clade IV (<50% to 68%), and analyses excluding the long-branch rhabditid taxa *Bunonema*, *Teratorhabditis* and *Pellioiditis* gave increased support for the Diplogasterida–Rhabditina clade V (51% to 93%). Figure 2a shows a consensus of these analyses: branchpoints that were supported by >60% in bootstrap long-branch taxa resampling of NJ or MP trees from the trimmed datasets were accepted. When there was no support for a resolved branching order, we collapsed nodes to form polytomies. Major clades supported by all analytical methods are shown and are numbered I–V. The association of Secernentea (clades II, IV and V) with the Plectidae has not been unequivocally resolved and is shown as a polytomy. We could not place the long-branch-length taxa in our trees with any certainty. We assessed statistical support for the placement of the vertebrate-parasitic taxa into four clades by calculating ML values for six-taxon subsets from the data. Each of the placements was strongly supported.

Received 16 July; accepted 12 December 1997.

- Luc, M., Sikora, R. A. & Bridge, J. *Plant Parasitic Nematodes in Tropical and Subtropical Agriculture* (CAB International, Wallingford, UK, 1990).
- Anderson, R. C. *Nematode Parasites of Vertebrates. Their Development and Transmission* (CAB International, Wallingford, UK, 1992).
- Lambhead, J. Recent developments in marine benthic biodiversity research. *Oceanis* **19**, 5–24 (1993).
- Riddle, D., Blumenthal, T., Meyer, B. & Priess, J. (eds) *C. elegans II* (Cold Spring Harbor Laboratory Press, NY, 1997).
- Ellis, R. E., Sulston, J. E. & Coulson, A. R. The rDNA of *C. elegans*: sequence and structure. *Nucleic Acids Res.* **14**, 2345–2364 (1986).
- Zarlenga, D. S., Stringfellow, E., Nobary, M. & Lichtenfels, J. R. Cloning and characterisation of ribosomal RNA genes from three species of *Haemonchus* (Nematoda: Trichostrongyloidea) and identification of PCR primers for rapid differentiation. *Exp. Parasitol.* **78**, 28–36 (1994).
- Fitch, D. H. A., Bugaj-gaweda, B. & Emmons, S. W. 18S ribosomal gene phylogeny for some rhabditidae related to *Caenorhabditis elegans*. *Mol. Biol. Evol.* **12**, 346–358 (1995).
- Baldwin, J. G., Frisse, L. M., Vida, J. T., Eddleman, C. D. & Thomas, W. K. An evolutionary framework for the study of developmental evolution in a set of nematodes related to *Caenorhabditis elegans*. *Mol. Phylogenet. Evol.* **8**, 249–259 (1997).
- Baldwin, J. G. *et al.* The buccal capsule of *Aduncoscipulum halicti* (Nemata: Diplogasterina): an ultrastructural and molecular phylogenetic study. *Can. J. Zool.* **75**, 407–423 (1997).
- Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. in *Molecular Systematics* (eds Hillis, D. M., Moritz, C. & Mable, B. K.) 407–514 (Sinauer, Sunderland, MA, 1996).
- Aguiñaldo, A. M. A. *et al.* Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**, 489–493 (1997).
- Lorenzen, S. *The Phylogenetic Systematics of Free-Living Nematodes* (The Ray Society, London, 1994).
- Malakhov, V. V. *Nematodes. Structure, Development, Classification and Phylogeny* (Smithsonian Institution Press, Washington, 1994).
- Maggenti, A. R. in *Concepts in Nematode Systematics* (eds Stone, A. R., Platt, H. M. & Khalil, L. F.) 25–40 (Academic, London, 1983).
- Baldwin, J. G. & Eddleman, C. D. Buccal capsule of *Zeldia punctata* (Nemata: Cephalobidae): an ultrastructural study. *Can. J. Zool.* **73**, 648–656 (1995).
- Etlinger, A. & Sommer, R. The homeotic gene *lin-39* and the evolution of nematode epidermal cell fates. *Science* **278**, 452–455 (1997).
- Poinar, G. Origins and phylogenetic relationships of the entomophilic rhabditids, *Heterorhabditis* and *Steinernema*. *Fund. Appl. Nematol.* **16**, 332–338 (1993).
- Siddiqi, M. R. Phylogenetic relationships of the soil orders Dorylaimida, Mononchida, Triplonchida and Alaimida, with a revised classification of the subclass Enoplia. *Pak. J. Nematol.* **1**, 79–110 (1983).
- Poinar, G. O. *The Natural History of Nematodes* (Prentice-Hall, Englewood Cliffs, NJ, 1983).
- De Ley, P., van de Velde, M. C., Mounport, D., Baujard, P. & Coomans, A. Ultrastructure of the stoma in Cephalobidae, Panagrolaimidae and Rhabditidae, with a proposal for a revised stoma terminology in Rhabditidae. *Nematologica* **41**, 153–182 (1995).
- Winnepernick, B. *et al.* 18S rRNA data indicate that Aschelminthes are polyphyletic in origin and consist of at least three distinct clades. *Mol. Biol. Evol.* **12**, 1132–1137 (1995).
- Blaxter, M. L. *et al.* Genes expressed in *Brugia malayi* infective third stage larvae. *Mol. Biochem. Parasitol.* **77**, 77–96 (1996).
- Swofford, D. L. *PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1* (Illinois Natural History Society, Champaign, 1993).
- Maddison, W. & Maddison, D. *MacClade v3.0* (Sinauer, Sunderland, MA, 1993).
- Kumar, S., Tamura, K. & Nei, M. *MEGA: Molecular Evolutionary Genetics Analysis. Version 1.0* (Pennsylvania State Univ., 1993).

- Van de Peer, Y., Rensing, S., Maire, U.-G. & De Wachter, R. Substitution rate calibration of small subunit subunit RNA identifies chlorarachniophyte nucleomorphs as remnants of green algae. *Proc. Natl Acad. Sci. USA* **93**, 7732–7736 (1996).
- Van de Peer, Y. & De Wachter, R. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput. Appl. Biosci.* **10**, 569–570 (1994).
- Yang, Z. *Phylogenetic Analysis by Maximum Likelihood (PAML) Version 1.2* (Univ. California, Berkeley, 1996).
- Felsenstein, J. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410 (1978).

Acknowledgements. We thank our colleagues for donations of nematode material, and D. Swofford for use of prerelease versions of PAUP*4.0. This work was supported by grants from the Wellcome Trust, the Linnean Society of London, the Belgian National Fund for Scientific Research, the NSF, the NIH and the United States Department of Agriculture.

Correspondence and requests for materials should be addressed to M.L.B. (e-mail: mark.blaxter@ed.ac.uk).

Spatial and temporal organization during cardiac fibrillation

Richard A. Gray[†], Arkady M. Pertsov^{*} & José Jalife^{*}

^{*} Department of Pharmacology, SUNY Health Science Center, 766 Irving Avenue, Syracuse, New York 13210, USA

[†] Department of Biomedical Engineering and Department of Medicine Division of Cardiovascular Disease, University of Alabama at Birmingham, 1670 University Blvd, Birmingham, Alabama 35294-0019, USA

Cardiac fibrillation (spontaneous, asynchronous contractions of cardiac muscle fibres) is the leading cause of death in the industrialized world¹, yet it is not clear how it occurs. It has been debated whether or not fibrillation is a random phenomenon. There is some determinism during fibrillation^{2,3}, perhaps resulting from rotating waves of electrical activity^{4–6}. Here we present a new algorithm that markedly reduces the amount of data required to depict the complex spatiotemporal patterns of fibrillation. We use a potentiometric dye⁷ and video imaging^{8,9} to record the dynamics of transmembrane potentials at many sites during

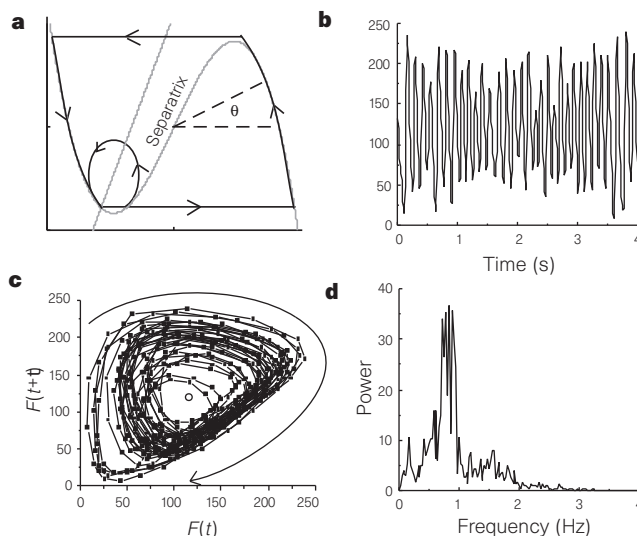


Figure 1 Temporal organization. **a**, Phase portrait of an excitable element incorporating two state variables³⁰. A stable fixed point occurs at the intersection of the nullclines (dotted lines)³⁰. **b**, Fluorescence signal (F) from a site on the surface of a rabbit heart during fibrillation. **c**, Phase portrait reveals trajectories circling around a centre (F_{mean} , F_{mean}), shown as a circle. **d**, The fluorescence signal exhibited a periodic component centred near 8 Hz, as observed in the corresponding power spectra. The frequency band 8 ± 3 Hz was different from equivalent white noise; $P < 0.00001$ for each heart (all sites combined).

fibrillation. Transmembrane signals at each site exhibit a strong periodic component centred near 8 Hz. This periodicity is seen as an attractor in two-dimensional phase space and each site can be represented by its phase around the attractor. Spatial phase maps at each instant reveal the 'sources' of fibrillation in the form of topological defects, or phase singularities¹⁰, at a few sites. Using our method of identifying phase singularities, we can elucidate the mechanisms for the formation and termination of these singularities, and represent an episode of fibrillation by locating singularities. Our results indicate an unprecedented amount of temporal and spatial organization during cardiac fibrillation.

It is still uncertain whether rotors underlie cardiac fibrillation. Self-organized rotors giving rise to spiral waves have been observed in various excitable media^{11–13} including cardiac muscle⁸. Although stationary spiral waves occur in isolated thin pieces of cardiac tissue, in the whole heart, as in many excitable media, they tend to move throughout the heart. If these spiral waves move rapidly (at >30% of the wave speed), they give rise to fibrillatory activity⁴. The mechanisms of cardiac fibrillation vary^{4,15}, however, and fibrillation is usually the result of multiple three-dimensional electrical waves, sometimes described as meandering wavelets, propagating throughout the heart^{16,17}. Cardiac fibrillation has been described in terms of

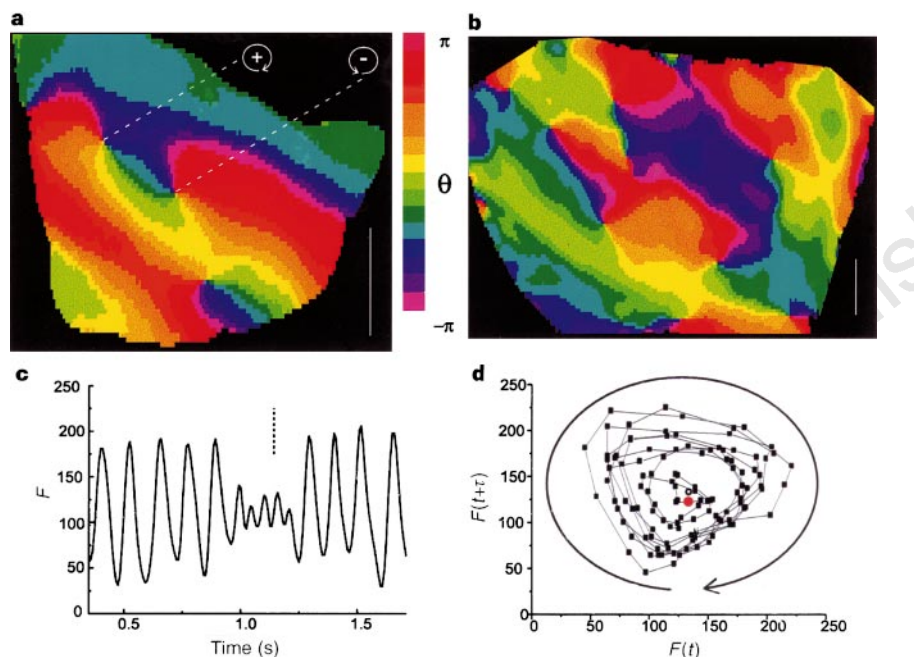


Figure 2 Snapshots of phase from the heart surface of the rabbit and sheep during sustained fibrillation. **a**, Rabbit; **b**, sheep. We classify rotor chirality as '+' for clockwise and '-' for anticlockwise²⁵. At these instants, three phase singularities (two clockwise and one anticlockwise) were observed on the rabbit heart and nine (five clockwise and four anticlockwise) on the larger sheep heart. Signals (F) demonstrate **(c)** low amplitude and **(d)** remain near the centre of their phase portraits when a spatial phase singularity site is nearby. Dashed line and red circle indicate the time of the corresponding snapshot. Vertical white line represents 1 cm.

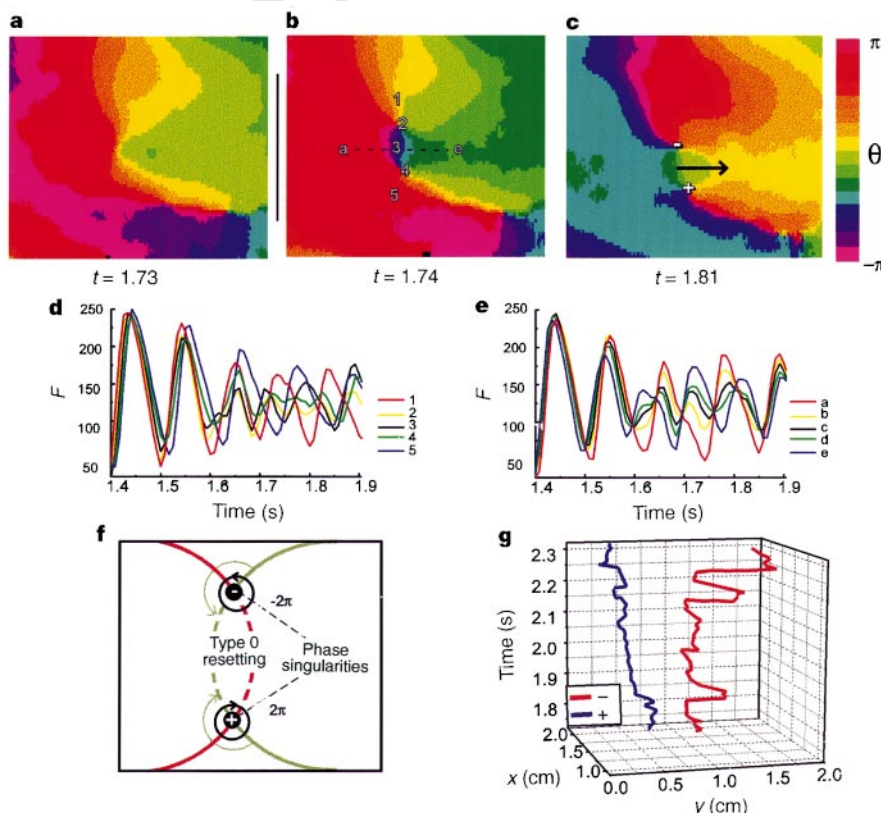


Figure 3 Initiation of a pair of spatial phase singularities. Snapshots of phase before **(a)**, during **(b)**, and after **(c)** the formation of a pair of spatial phase singularities during sustained fibrillation in the sheep heart. **d**, **e**, Transmembrane signals (F) measured at sites **a–e** and **1–5** labelled in **b**, **f**. A pair of singularities form when the local phase gradient becomes large (in other words, the excitation wave, $\theta \approx 0$ (green), approaches regions not fully recovered, $\theta \approx \pm \pi$ (red)). The excitation wave cannot proceed into the recovered region, and hence breaks, forming two phase singularities. The two excitation waves rotate around these newly formed singularities. Sustained rotation in the form of a pair of rotors occurs only if this excitation wave causes type 0, or even, phase resetting at the site of the initial wave break. Type 0 resetting (suprathreshold) advances the phase of this region into a new cycle, generating a new excitation wave (in the opposite direction to the previous wave; see arrow in **c**), resulting in the formation of a pair of self-sustaining rotors. Type 1 resetting (subthreshold) does not create this new excitation wave, and the phase singularity pair lasts less than one rotation. Notice the 'extra' cycle in the central region of block, sites **3** & **c** in **d** and **e**, indicative of type 0 resetting. **g**, Trajectories of '+' and '-' rotors following their initiation plotted in x , y , t space. Vertical line between **a** and **b** represents 1 cm.

rotors^{4–6,14,15}, on the basis of the long-held view that the heart is an example of a generic excitable medium^{10,18–20}. There are many theories about fibrillation in excitable media, but only recently have experimental techniques become available to study the complex spatial patterns observed during sustained fibrillation^{4,6,21–23}. Results from these experiments indicate that rotating waves are observed during fibrillation; however, they appear infrequently, and their initiation, termination and interaction have not been characterized.

An excitable element (for example, a cell, a patch of membrane, or a localized region in a spatially distributed system) can often be represented in phase space, which the element spends most of its time at a fixed point. A suprathreshold stimulus pushes the state of the excitable element past the separatrix and it continues along a closed-loop trajectory; however, if the stimulus is subthreshold, the state of the element does not cross the separatrix (Fig. 1). In periodic dynamics, it is simple and useful to represent the state of an element by its phase (θ) around the loop. The responses of single elements to external stimuli have been extensively studied by analysing the induced changes in θ (that is, phase resetting). Two fundamentally different responses to stimuli occur, namely, type 0 or even resetting, where a suprathreshold response gives rise to a new cycle, and type 1 or odd resetting, which effects a subthreshold response to stimuli^{10,24}. In spatially extended excitable systems, the stimulation of individual elements are provided by neighbouring elements (usually through diffusion).

A rotor is composed of a wave of excitation propagating around a topological defect, which is known as a phase singularity. A spatial phase singularity is a site in an excitable medium at which the phase of the site is arbitrary; the neighbouring elements exhibit a continuous progression of phase that is equal to $\pm 2\pi$ around this site. As shown in Fig. 1, transmembrane signals (F) recorded from the surface of rabbit and sheep hearts during fibrillation exhibited attractors in reconstructed two-dimensional phase space, when $F(t)$ was plotted against $F(t + \tau)$ where t is time and τ is the embedding delay (see Methods). Periodicity of each site was near 8 Hz. Although trajectories for subsequent cycles did not coincide in

phase space, the trajectories circulated around a central region, allowing us to construct a new variable, the phase along the attractor, θ . The phase variable (θ) calculated at each site rotated clockwise ($d\theta < 0$) 89 $\pm$ 4% of the time, indicating that the trajectories were encircling the centre. The distance of points from the centre was smaller for $d\theta \geq 0$ compared with $d\theta < 0$ ($P < 0.0001$), as would be expected if the phase were ambiguous at the centre¹⁰. This new variable, θ , has certain advantages over the fluorescence signal (F) that simplify the analysis of fibrillation. First, use of θ eliminates the need to pick activation times, which is difficult during fibrillation, especially in the important regions of slow propagation and block. Second, we can test directly whether spatial phase singularities exist and are necessary to maintain fibrillation.

With this new method to represent fibrillation by phase, $\theta(x, y, t)$, we could study directly the detailed dynamics of spatial phase singularities and rotors, including their initiation and termination. The spatial phase patterns during sustained fibrillation (Fig. 2) concurred with theoretical predictions (for example, isophase lines connect phase singularities of opposite chirality or end on a no-flux boundary)^{10,25}. Spatial phase singularities are easily identified as sites at which all phase values ($-\pi$ to π) converge. The continuous spatial phase changes reflect waves propagating on the heart surface as a result of processes of excitation, recovery, and diffusion, and indicate that each site of the heart surface can be represented by its phase around a two-dimensional attractor. We elucidated the mechanism of rotor formation and termination by analysing successive frames of $\theta(x, y, t)$, (representative examples are shown in Figs 3 and 4). Movement of existing phase singularities created Doppler-shifted^{4,8,26} short cycle lengths, and thus created large local phase gradients in front of moving singularities. The formation of phase singularities was necessary, although not sufficient, for sustained rotation (that is, for rotor formation). In addition to the formation of phase singularities, a new excitation wave must be generated²⁷ (that is, type 0 or even resetting must occur)^{10,24} to form a rotor. Lifespan histograms for both rabbits and sheep indicate that the majority (80% for rabbit and 84% for sheep) of phase singularities lasted < 100 ms, which is less than one rotation^{4,14}. Therefore,

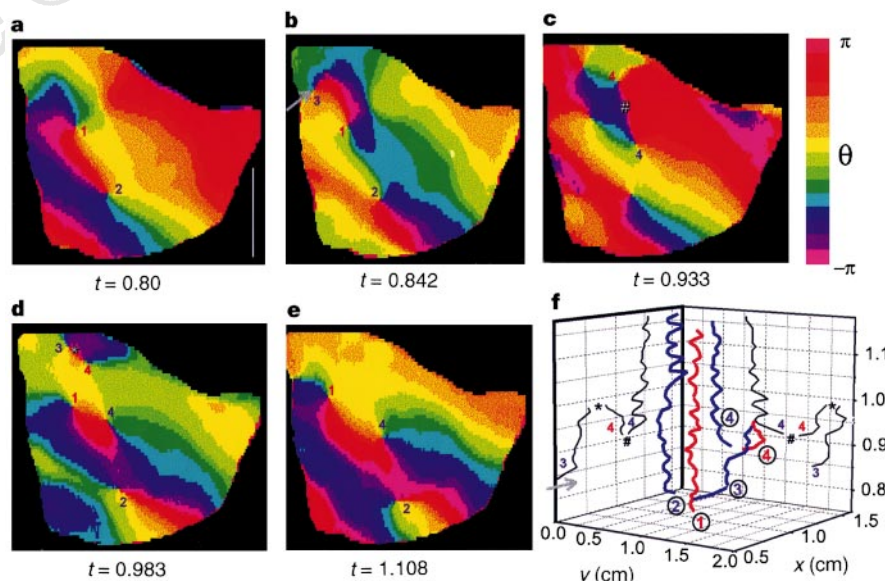


Figure 4 Dynamics of phase singularities. **a–e**, Snapshots of phase illustrating rotor dynamics on the surface of the rabbit heart during sustained fibrillation. **f**, Trajectories of the clockwise and anticlockwise rotors shown in **a–e** plotted in x, y, t space. Each rotor is numbered (1–4) and coloured (clockwise, blue; anticlockwise, red). The x and y projections in time are shown in black in **f** (except for rotors 1 and 2 for clarity). **a**, Rotors 1 and 2 formed separately before the time interval shown here (0.75–1.17 s). **b**, At $t = 0.842$ s a clockwise rotor (3) enters the

field of view from the left and moves rapidly toward the right (grey arrow in **b** and **f**). This movement creates a convergence of phase values ahead of rotor 3, and (**c**) a pair of phase singularities (4) form (# in **c** and **f**) when the excitation wave ($\theta = 0$) reaches the high-phase-gradient region. Both of these newly formed singularities move; the clockwise one collides with anticlockwise-rotating rotor 3, resulting in mutual annihilation (asterisk in **d**), whereas the anticlockwise one survives in the form of a rotor. Vertical white line represents 1 cm.

only ~20% of phase singularities formed rotors.

Rotor termination occurred when rotors of opposite chirality merged or when a single rotor collided with a boundary (both of these occurrences are topologically equivalent for a non-flux boundary)¹⁰. Specifically, rotor pairs were mutually annihilated if the phase gradient between the rotors, perpendicular to the line connecting the phase singularities, was sufficiently large to stop propagation. On the basis of the number of rotors observed in our recording array, each rotor occupied on average $12 \pm 4 \text{ cm}^2$. According to rough measurements of heart surface area, we estimate that the total number of rotors during fibrillation would be approximately 1–2 for rabbits, 5 for sheep, and 15 for humans (assuming the rotor density is the same in humans).

These results indicate that analysing the complex spatiotemporal patterns seen during fibrillation on the surface of the heart can be greatly simplified by identifying and analysing phase singularities. This analysis reveals topological restrictions to the dynamics of fibrillation¹⁰: first, phase lines do not intersect; second, phase singularities are joined via isophase lines to two other singularities with opposite chirality (or a boundary); and third, phase singularities form and terminate as oppositely rotating pairs (Fig. 4). Under certain conditions, phase singularities give rise to rotors, which sustain fibrillation. The direct observation of phase singularities has led, for the first time, to the quantification of fibrillation in terms of rotor number and lifespan, and to the elucidation of the mechanisms underlying the formation and termination of rotors. □

Methods

The experimental protocols, video imaging-recording system, and signal processing have been described previously^{14,28}. We acquired video images (typically 200×100 pixels) from the ventricular surface at a rate of 120 frames s^{-1} ($\Delta t = 0.00833 \text{ s}$). We applied spatial and temporal filtering to improve the signal-to-noise ratio²⁸. Two-dimensional phase portraits were obtained by plotting $F(t)$ versus $F(t + \tau)$ (ref. 29), where $t + n\Delta t$ and n is the frame number. The value of τ was chosen to be roughly one-quarter of the cycle length during fibrillation ($\tau = 25 \text{ ms}$); this value roughly corresponds to the first zero crossing of the autocorrelation of F_i indicating linear independence. A new variable phase, $\theta(t)$, was computed as $\text{atan}(F(t + \tau) - F_{\text{mean}}, F(t) - F_{\text{mean}})$. Isolated hearts from rabbits of ~3 kg ($n = 3$) and from sheep of ~20 kg ($n = 3$) body weight were maintained at $36\text{--}38^\circ\text{C}$. Ventricular fibrillation was initiated by rapid pacing, and 4-s recordings were obtained at least 5 min after initiation (perfusion was maintained during fibrillation). In Figs 2–4, white and black bars near images reflect a distance of 1 cm. Values are presented as mean $\pm$ s.d.m. Comparisons were made with paired student t -tests.

Received 30 September; accepted 8 December 1997.

- Myerburg, R. J. *et al.* in *Cardiac Electrophysiology, From Cell to Bedside* (W. B. Saunders, Philadelphia, 1990).
- Garfinkel, A. *et al.* Quasiperiodicity and chaos in cardiac fibrillation. *J. Clin. Invest.* **99**, 305–314 (1997).
- Witkowski, F. X. *et al.* Evidence for determinism in ventricular fibrillation. *Phys. Rev. Lett.* **75**, 1230–1233 (1995).
- Gray, R. A. *et al.* Mechanisms of cardiac fibrillation. *Science* **270**, 1222–1223 (1995).
- Ikeda, T. *et al.* Mechanism of spontaneous termination of functional reentry in isolated canine right atrium. *Circulation* **94**, 1962–1973 (1996).
- Lee, J. J. *et al.* Reentrant wavefronts in Wiggers' stage II ventricular fibrillation. *Circ. Res.* **78**, 660–675 (1996).
- Salzberg, B. M., Davila, H. V. & Cohen, L. B. Optical recordings of impulses in individual neurons of an invertebrate central nervous system. *Nature* **246**, 508–509 (1973).
- Davidenko, J. M., Pertsov, A. M., Salomonsz, R., Baxter, W. T. & Jalife, J. Stationary and drifting spiral waves of excitation in isolated cardiac muscle. *Nature* **355**, 349–351 (1991).
- Pertsov, A. M., Davidenko, J. M., Salomonsz, R., Baxter, W. T. & Jalife, J. Spiral waves of excitation underlie reentrant activity in isolated cardiac muscle. *Circ. Res.* **72**, 631–650 (1993).
- Winfree, A. T. *When Time Breaks Down*. (Princeton Univ. Press, 1987).
- Winfree, A. T. Scroll-shaped waves of chemical activity in three dimensions. *Science* **181**, 937–939 (1973).
- Goldbeter, A. Mechanism for oscillatory synthesis of cAMP in *Dictyostelium discoideum*. *Nature* **253**, 540–542 (1975).
- Lechleiter, J., Girdard, S. & Peralta, E. Spiral calcium wave propagation and annihilation in *Xenopus laevis* oocytes. *Science* **252**, 123–125 (1991).
- Gray, R. A. *et al.* Non-stationary vortex-like reentry as a mechanism of polymorphic ventricular tachycardia in the isolated rabbit heart. *Circulation* **91**, 2454–2469 (1995).
- Winfree, A. T. Electrical turbulence in three-dimensional heart muscle. *Science* **266**, 1003–1006 (1994).
- Moe, G. K. & Abildskov, J. A. Atrial fibrillation as a self-sustaining arrhythmia independent of focal discharge. *Am. Heart J.* **58**, 59–70 (1959).

- Allessie, M. A., Lammers, W., Bonke, F. I. M. & Hollen, J. in *Cardiac Electrophysiology and Arrhythmias* 265–275 (Grune and Stratton, Orlando, 1985).
- Krinsky, V. I. Mathematical models of cardiac arrhythmias (spiral waves). *Pharmacol. Ther.* **B3**, 539–555 (1978).
- Zykov, V. S. *Simulation of Wave Processes in Excitable Media*. (University Press, New York/Manchester, 1987).
- Gray, R. A. & Jalife, J. Spiral waves and the heart. *Int. J. Bifurc. Chaos* **6**, 415–435 (1996).
- Bayly, P. V. *et al.* Efficient electrode spacing for examining spatial organization during ventricular fibrillation. *J. Cardiovasc. Electrophysiol.* **4**, 533–546 (1993).
- Bove, R. T. & Dillon, S. M. A new high performance system for imaging cardiac electrical activity. *Circulation* **94**, 1–714 (1996).
- Cha, Y. M., Birgersdotter-Green, U., Wolf, P. L., Peters, B. B. & Chen, P. S. The mechanism of termination of reentrant activity in ventricular fibrillation. *Circ. Res.* **74**, 495–506 (1994).
- Glass, L. & Mackay, M. C. *From Clocks to Chaos* (Princeton Univ. Press, 1988).
- Agladze, K. I. & Krinsky, V. I. Multi-armed vortices in an active chemical medium. *Nature* **296**, 424–426 (1982).
- Fast, V. G. & Pertsov, A. M. Drift of a vortex in the myocardium. *Biophysics* **35**, 489–494 (1990).
- Gotoh, M. *et al.* Cellular graded responses and ventricular vulnerability to reentry by a premature stimulus in isolated canine ventricle. *Circulation* **95**, 2141–2154 (1997).
- Gray, R. A., Ayers, G. & Jalife, J. Video imaging of atrial defibrillation in the sheep heart. *Circulation* **95**, 1038–1047 (1997).
- Takens, F. in *Dynamical Systems and Turbulence* (eds Rand, D. A. & Young, L. S.) *Lecture Notes in Mathematics* **898**, 366–381 (Springer, Berlin, 1981).
- Fitzhugh, R. Impulses and physiological states in theoretical models of nerve membrane. *Biophys. J.* **1**, 445–446 (1961).

Acknowledgements. We thank O. Berenfield, Z. Silverman, J. Jiang and M. Flanagan for technical assistance and M. Vinson for reading the manuscript. This work was supported by grants from the Whitaker Foundation and the N.I.H.

Correspondence and requests for materials should be addressed to R.A.G. (e-mail: rag@crml.uab.edu).

Spatiotemporal evolution of ventricular fibrillation

Francis X. Witkowski*, L. Joshua Leon†, Patricia A. Penkoske‡, Wayne R. Giles§, Mark L. Spano||, William L. Ditto¶ & Arthur T. Winfree#

* Department of Medicine, and ‡ Department of Surgery, University of Alberta, Edmonton, Alberta T6G 2R7, Canada

† Ecole Polytechnique, Montreal, Quebec H3C 3J7, Canada

§ Department of Physiology and Biophysics, University of Calgary, Calgary, Alberta T2N 4N1, Canada

|| Naval Surface Warfare Center, West Bethesda, Maryland 20817, USA

¶ Applied Chaos Laboratory, School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA

Sudden cardiac death is the leading cause of death in the industrialized world, with the majority of such tragedies being due to ventricular fibrillation¹. Ventricular fibrillation is a frenzied and irregular disturbance of the heart rhythm that quickly renders the heart incapable of sustaining life. Rotors, electrophysiological structures that emit rotating spiral waves, occur in several systems that all share with the heart the functional properties of excitability and refractoriness. These re-entrant waves, seen in numerical solutions of simplified models of cardiac tissue², may occur during ventricular tachycardias^{3,4}. It has been difficult to detect such forms of re-entry in fibrillating mammalian ventricles^{5–8}. Here we show that, in isolated perfused dog hearts, high spatial and temporal resolution mapping of optical transmembrane potentials can easily detect transiently erupting rotors during the early phase of ventricular fibrillation. This activity is characterized by a relatively high spatiotemporal cross-correlation. During this early fibrillatory interval, frequent wavefront collisions and wavebreak generation⁹ are also dominant features. Interestingly, this spatiotemporal pattern undergoes an evolution to a less highly spatially correlated mechanism that lacks the epicardial manifestations of rotors despite continued myocardial perfusion.

Ventricular fibrillation is a complicated, often lethal, but poorly understood, high-frequency mode of electrical activity. It can be

initiated abruptly in both normal and compromised heart muscle and can, just as abruptly, be electrically terminated¹⁰. Ventricular fibrillation can exist as a stable state which is self-sustained and independent of its initiating event. The mechanisms underlying the pernicious stability of ventricular fibrillation are unknown, and explanations of how shocks terminate ventricular fibrillation are speculative. Present methods of terminating ventricular fibrillation are open loop: a large shock is applied to the fibrillating heart, which is then left to recover and resume normal spontaneous beating. This is the only known therapy for ventricular fibrillation. Patients suffering recurrent ventricular fibrillation are often treated by implantation of an automatic defibrillation system.

In 1930, Carl Wiggers used a movie camera operating at 32 frames per second to record the movements of the surface of the *in situ* heart in which ventricular fibrillation was induced electrically¹¹. He described four stages: first, an initial phenomenon consisting of 2–8 rapidly activated peristaltic waves; second, a subsequent convulsive incoordination stage that lasted 14–40 seconds, so named because “When the ventricles are held in the palm of the hand, a fluttering, undulatory, convulsive sensation is experienced” although the heart cannot generate any blood pressure; third and fourth, subsequent stages that reflected the progressive ischaemia. In terms of clinical interventions, the most significant of these stages is Wiggers’ stage 2, when appropriate countermeasures can be instituted and sudden death aborted. Using an electronic camera system operating at 60 frames per second together with staining of the heart with voltage-

sensitive dye, a single rapidly moving rotor that produced an electrocardiographic pattern that resembled fibrillation has been described in rabbit hearts⁵. Searches for similar rotating waves in larger mammalian hearts have identified spiral waves in sheep heart during ventricular fibrillation⁸. However, these are described as uncommon occurrences in canine hearts⁶ and rare in porcine hearts⁷.

We recorded the electrical activity from a limited epicardial area (~30%) of the right and left ventricles in isolated blood-perfused canine hearts (Fig. 1). The device used consists of an image-intensified charge-coupled device (CCD) optical recording system¹², used to image an epicardial surface stained with the voltage-sensitive dye di-4-ANEPPS. Fluorescent images were obtained with a temporal resolution of 1.2 ms and a spatial resolution of approximately 0.5 mm (see Methods and ref. 12 for details). The degree of spatial correlation in sequential images was quantified using a peak cross-correlation method as shown in Fig. 2, based on the rationale that repeating patterns of activity would have higher peak cross-correlations. Ventricular fibrillation was induced with a single critically timed electrical pulse (S2), applied after a constant pacing train (of S1 pulses). The heart was defibrillated with a biphasic shock from an external defibrillator, applied by paddles placed on the right and left ventricles. Both the early onset of ventricular fibrillation (corresponding to Wiggers’ stages 1 and 2) and sustained ventricular fibrillation that lasted for more than 10 minutes in perfused hearts were imaged with this apparatus. The optical transmembrane potentials and their temporal derivatives were then viewed as movies (Quicktime and AVI movie sequences for this study are available as Supplementary Information).

An example of the appearance of normal (S1-induced) activation wavefronts is given in Fig. 3c; this shows the familiar anisotropy associated with fibre orientation. The sequential data from 1,000 frames of optical images that include two such S1 stimuli and the induction of ventricular fibrillation by the S2 stimulus are shown in Fig. 3e. The initial waves triggered by the S2 consistently produced a re-entrant cycle with a ‘figure-of-eight’ morphology¹³; this cycle was composed of two mirror image rotors that shared a common

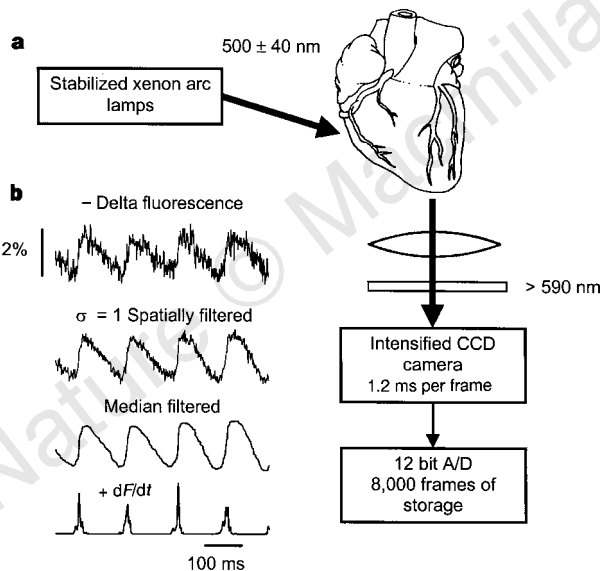


Figure 1 The procedure used to record the electrical activity in blood-perfused canine hearts. **a**, Schematic representation of the experimental set-up. Two 1-kW, stabilized, xenon arc lamps were used to illuminate the epicardial cardiac surface under study with ~ 100 mW per cm^2 of quasi-monochromatic green light (wavelengths 500 ± 40 nm). This focused fluorescent source was imaged after barrier-filter rejection of reflected light components and retention of fluorescent components with wavelengths > 590 nm, as illustrated. The cooled, fibre-optically coupled, image-intensified CCD frame-transfer camera system was operated at 1.19 ms per frame with 12 bits of dynamic range. **b**, Time series from the sequential processing steps for a representative single pixel from 500 frames during ventricular fibrillation. The effects (on signal to noise) of subsequent image processing steps are depicted; these steps included 9×9 gaussian spatial averaging (with standard deviation $\sigma = 1$ for this kernel, being the radius in pixels containing 68% of the integrated magnitude of the coefficients¹⁵), followed by 21-point median temporal filtering, and culminating with 5-point estimation of temporal derivatives with final clipping to maintain only the positive values (setting all negative values to 0). The optical calibration bar indicates a 2% change in fluorescence.

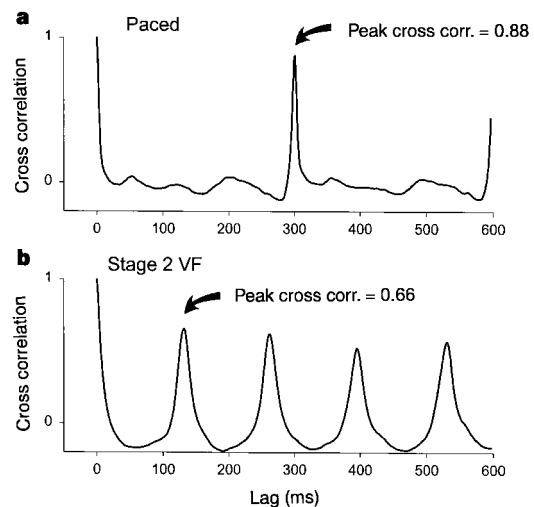


Figure 2 Quantification of the degree of spatial correlation in sequential images. **a**, Cross-correlation as a function of lag (elapsed time) for a ventricular stimulus applied at a basic cycle length of 300 ms. Peak cross-correlation (Peak cross corr.), expressed as a function of the selected lag interval (50–300 frames) for this index ensemble of 41 frames of $+dF/dt$ images, occurs at a frame lag corresponding to the pacing interval. **b**, Cross-correlation calculated during ventricular fibrillation (VF) after the initial ‘figure-of-eight’ re-entry induced by the S2 stimulation has terminated. Peak value is indicated by curved arrow. This electrophysiological response after induction of ventricular fibrillation is commonly referred to as Wiggers’ stage 2.

re-entrant pathway. The time of onset of the S2 and the duration of the re-entry which it directly initiated are colour coded with green shading in the temporal stack (Fig. 3e). The corresponding peak cross-correlation relationship (computed from these 1,000 frames of data) is shown in Fig. 3f. There is a high peak cross-correlation for the first of the S1 stimulations, indicating the similar appearance of the transmembrane potential information induced by the second of the S1 stimulations. The peak cross-correlation is reduced following the second stimulus in the train, as it is quickly followed by an activation pattern that appears quite different, that is, the pattern initiated by the S2 stimulus. In this example, the S2-induced re-entry lasted for a total of eight cycles before being abolished by wavefronts that collided in the area of this initial re-entry. These collisions often resulted in the subsequent emergence of two oppositely directed, spatially discrete wavefronts (with observable dangling ends)¹². Each of the dangling ends of the emerging wavefronts is also called a wavebreak or phase singularity⁹. Available movie files of data (see Supplementary Information) illustrate these features. Thereafter, rotors formed as shown in Fig. 3d, g, h; however, these events were never long-lived or reproducible in location. Indeed, in one animal in which ventricular fibrillation was induced under very similar stimulation conditions, the patterns that evolved diverged within the first few re-entrant cycles induced by the S2 stimulus. All episodes of induced ventricular fibrillation

were self-sustaining and terminated only when the heart was defibrillated.

A completely different electrophysiological pattern was seen when perfused ventricular fibrillation had persisted for 10 minutes. During 'chronic' ventricular fibrillation, no rotors were observed (see the movie files, Supplementary Information). Additionally, the peak cross-correlation (Fig. 4) was markedly reduced, although the period was roughly the same length as or slightly shorter than the pattern shown in Fig. 3g. The source mechanism is probably still re-entry, but its geometric aspect is now more three-dimensional. This pattern was reversible in that, after defibrillation and a recuperation interval of 10 minutes, the acute pattern of ventricular fibrillation (Wiggers' stage 2) was once again the initial manifestation after induction of ventricular fibrillation. Very similar waveform dynamics and numerical signatures were seen in six acute episodes of ventricular fibrillation (mean peak cross correlation = 0.50 ± 0.10) and four chronic perfused episodes of ventricular fibrillation (mean peak cross correlation = 0.28 ± 0.05), with the summary statistics significantly different (independent sample *t*-test, $P = 0.004$; SPSS 7.5 for Windows).

These studies were performed in normal perfused hearts. The mechanism(s) of this evolution of the morphology of ventricular fibrillation is unknown. Further studies aimed at understanding the role of the Purkinje fibre network, rotation of ventricular wall fibres,

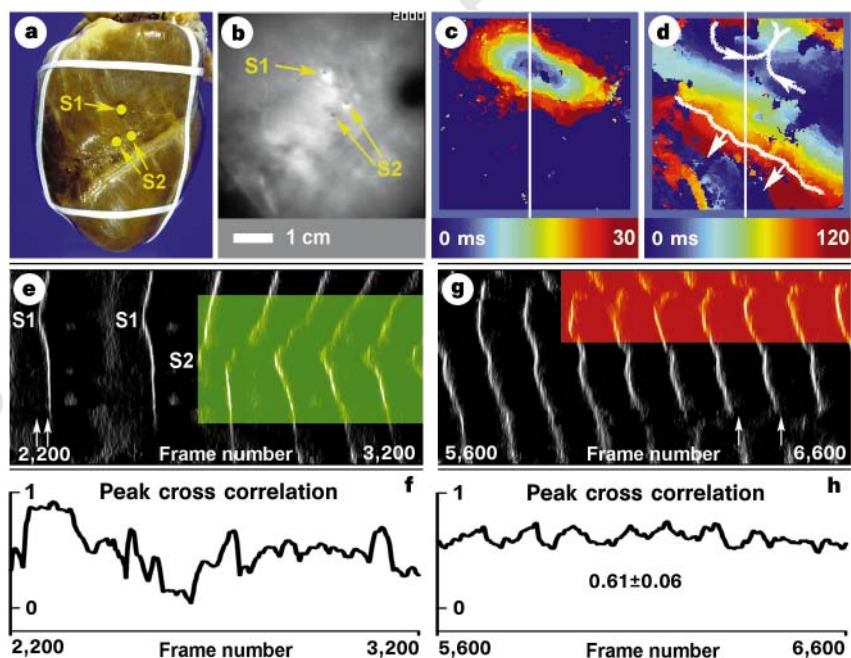


Figure 3 Overview of data observed with stimulated induction of ventricular fibrillation. **a**, Photograph of formalin-fixed canine heart with the pacing electrodes marked by yellow-headed pins, and the approximate area of epicardium that was imaged outlined with white tape. **b**, Overall fluorescence of epicardial surface with pacing sites on the right ventricle. At this magnification, each of the 128×128 pixels corresponds to a square of 0.43 mm on each side. **c**, Temporal derivative of optical transmembrane potentials for 25 frames during normal S1 pacing, with the $+dF/dt$ of each frame depicted using colours as indicated. This forms an image similar to images produced by activation isochrones, but without requiring any assumptions about what constitutes a local activation. The vertical white line indicates the column of pixels that was used to form the temporal stack of **e**. **d**, Colour-coded sequential plots of 100 frames of ventricular fibrillation data during rotor visualization within the field of view. Two rotors were present, as indicated by the white circular arcs, but only the anticlockwise rotor of the pair was well contained in the field of view. The two rotors' contributions were seen to merge and form the more linear wavefront as shown (white wavy line). Available data movie files (see Supplementary informa-

tion) illustrate these features. The vertical white line denotes the column of pixels used to form the temporal stack of the entire 1,000 images shown in **g**. **e**, Temporal stack of the $+dF/dt$ data due to two S1 stimuli followed by re-entrant activity induced by an extrastimulus, S2, highlighted by green shading. The temporal stack (1000×128 pixels) was produced by stacking one column of pixels (1×128 pixels) from 1,000 individual images edgewise. This initial phase of ventricular fibrillation corresponds to Wiggers' stage 1, and lasts for only eight re-entrant cycles in this example. The small vertical arrows indicate the frame locations for the 25 frames that were used to form **c**. **f**, Peak cross-correlations for the frames depicted in **e**, showing, first, the strong correlation between the first and second paced S1 beats, and second, the subsequent decrease in correlation as the response due to S2 occurs. **g**, Temporal stack for the frames indicated, with the region of the anticlockwise rotor formed highlighted by red shading. Vertical arrows indicate the first and last frames of the 100 frames used to form **d**. The mean cycle length is $\sim 130 \text{ ms}$. **h**, Peak cross-correlation corresponding to the frames shown in **g**, indicating an overall peak cross-correlation of 0.61 ± 0.06 (mean $\pm$ s.d.).

and wall thickness may all help to clarify this issue. With only the epicardial data presented, it is impossible to determine whether this evolution of morphology of ventricular fibrillation also involves the three-dimensional movement of rotors on the endocardial surface or in mid-myocardial regions. Mathematical modelling incorporating both realistic ventricular geometries and membrane ionic currents will assist in further investigations. Whether similar transformations occur in the progression of atrial fibrillation from paroxysmal to sustained fibrillation is interesting and can be studied using similar approaches. These insights into the spatiotemporal

characteristics of ventricular fibrillation should lead to improved therapy for fibrillation-induced sudden cardiac death, and provide an essential benchmark for mathematical simulation of ventricular arrhythmias. □

Methods

Imaging of blood-perfused isolated hearts and electrophysiology. Isolated blood-perfused canine (mongrel; weight range 17–21 kg, $n = 3$) hearts were prepared as described¹⁴. All animal procedures received approval from the Health Sciences Animal Welfare Committee of the University of Alberta. Studies were performed at 37 °C after the heart was stained with 5 μ M di-4-ANEPPS (Molecular Probes, Eugene, Oregon), infused via an aortic root cannula. The anterior right ventricle and part of the left ventricle (Fig. 3a) were compressed under an optical window to minimize motion artefacts. No pharmacological agents were used to reduce mechanical motion artefacts. An area of approximately 5.5 cm $\times$ 5.5 cm (30 cm²) was imaged; this represented ~30% of the epicardial surface.

Venous blood obtained before removal of the heart (~400 ml) was mixed with 1,500 ml perfusion medium containing (mM): Na⁺, 146.3; K⁺, 4.0; Cl⁻, 134.0; Ca²⁺, 1.0; Mg²⁺, 1.0; H₂PO₄, 0.3; HCO₃⁻, 20.0; and glucose, 20.0, with 400 units per litre heparin, 2 units per litre insulin, and dextran 40 g per litre. The total volume of this final blood-enriched perfusion fluid was recirculated in the perfusion apparatus after being gassed with 95% O₂/5% CO₂ by use of a capillary membrane oxygenator/heat exchanger. The hearts were perfused at a constant pressure of 75–80 mm Hg after excision and myocardial temperature was monitored with an intramyocardial temperature probe (Shiley TMPM, Shiley, Irvine, California). Ag/AgCl electrodes were sutured to the right and left ventricular free walls and provided DC-coupled biventricular electrogram recordings. We inserted pacing electrodes into the right ventricular free wall for S1 (unipolar, width 5 ms, twice diastolic threshold current) and S2 (bipolar, width 5–10 ms, 50–99 mA constant current) stimulation (Fig. 3a). Ventricular fibrillation was induced with a single extrastimulus (S2) applied after a train of 10–20 S1 pacing stimuli (with intervals of 300 ms between stimuli). Hearts were illuminated with quasi-monochromatic green light (500 $\pm$ 40 nm) from two 1-kW, short arc, stabilized xenon light sources (Model 66024 light source with 68850 light-intensity controller, Oriel, Stratford, CT), resulting in ~100 mW per cm² of illumination on the ventricular epicardial surface (Fig. 1). Focused fluorescent light was long-wavelength pass filtered (590 nm) before image intensification. At the illumination levels used, continuous recordings lasting 10–15 min could be realized with no detectable phototoxic damage.

We used a 128 $\times$ 128 pixel frame-transfer CA-D1-0128S CCD camera (Dalsa, Waterloo, Ontario, Canada) operating at 838 frames per second (1.19 ms per frame). It was cooled by 15 °C, and the analogue video signal underwent 12 bit analog-to-digital (A/D) conversion before transfer to an XPG-1000 frame grabber (Dipix Technologies, Ottawa, Ontario, Canada) with a peripheral computer interface (PCI) equipped with 256 MB of frame-grabber memory. This provided frame-grabber storage capacity for 8,000 frames of 128 $\times$ 128 pixels at 1.2 ms per frame, with 12 bits per pixel, which translates to 9.6 seconds of continuous real-time data acquisition. In addition, we constructed an isolated analogue data acquisition system consisting of 16 channels. These data were electronically deposited over a single line of the CCD output for simultaneous recording of stimulation train and biventricular electrograms. All software was custom-written. The read noise of our combined camera and A/D converter was measured to be 41.1 electrons (r.m.s.). Two-stage image intensification was incorporated between the barrier filter and the CCD camera, by using a single-microchannel-plate first-stage image intensifier (Gen III Ultra, ITT Night Vision, Roanoke, VA) whose output was fibre-optically coupled to a second-stage Gen I image intensifier (XX1490Z, DEP, Delft Instruments, Netherlands). The spatial resolution of the final image-intensified camera system was dictated by the resolution of the CCD imaging chip of 31.25 line pairs per mm, as verified using the 1951 USAF optical test pattern (Edmund Scientific, Barrington, NJ). After an initial, wavelet-based, spatially adaptive filter identified regions with viable signals from the CCD pixels, gaussian spatial filtering¹⁵ (s.d., 1 pixel) on the baseline-corrected individual pixel data, followed by 21-point temporal median filtering, was used, culminating with 5-point determination of temporal derivatives (Fig. 1b).

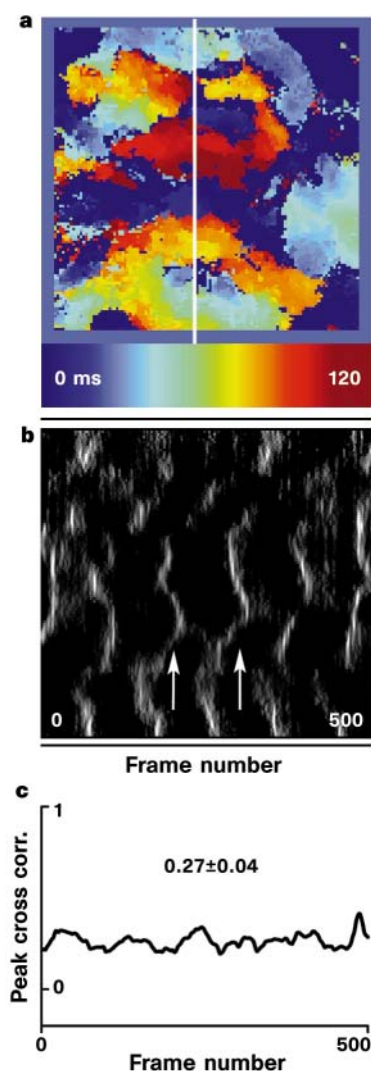


Figure 4 Electrophysiological patterns seen in chronic ventricular fibrillation. **a**, Colour-coded sequential $+dF/dt$ plots of 100 frames recorded during chronic perfused ventricular fibrillation. The patterns now seen are best described as wandering wavelets. The accompanying video movies of these fluorescent images (see Supplementary Information) show the dynamics and the striking differences in this form of ventricular fibrillation, compared with the patterns depicted in Fig. 3. The vertical white line denotes the column of pixels used to form the temporal stack shown in **b**. **b**, Temporal stack of the $+dF/dt$ data shows the progression of chronic ventricular fibrillation, with a decrease in organization as compared with Fig. 3g. The small vertical arrows indicate the frame locations for the 100 frames that were used to form **a**. The mean cycle length is ~115 ms. **c**, The peak cross-correlations for the frames depicted in **b** confirm in quantitative terms the visual impression that the chronic and acute forms of ventricular fibrillation differ in their spatiotemporal organization. The mean $\pm$ s.d. for the peak cross-correlation for the 500 frames analysed was now 0.27 ± 0.04 .

Received 26 November 1997; accepted 21 January 1998.

1. Myerburg, R. J. *et al.* in *Cardiac Electrophysiology, From Cell to Bedside* (eds Zipes, D. P. & Jalife, J.) 666–678 (W. B. Saunders, Philadelphia, 1990).
2. Holden, A. V. The restless heart of a spiral. *Nature* **387**, 655–657 (1997).
3. Winfree, A. T. Electrical turbulence in three-dimensional heart muscle. *Science* **266**, 1003–1006 (1994).
4. Panfilov, A. V. & Holden, A. V. (eds) *Computational Biology of the Heart* (Wiley, Chichester, 1997).
5. Gray, R. A. *et al.* Mechanisms of cardiac fibrillation. *Science* **270**, 1222–1225 (1995).
6. Lee, J. J. *et al.* Reentrant wave fronts in Wiggers' stage II ventricular fibrillation: characteristics and mechanisms of termination and spontaneous regeneration. *Circ. Res.* **78**, 660–675 (1996).
7. Rogers, J. M., Huang, J., KenKnight, B. H., Smith, W. M. & Ideker, R. E. Stationary reentrant circuits are rare during ventricular fibrillation in pigs. *Circulation* **94**, 1–48 (1996).
8. Gray, R. A. & Jalife, J. Self-organized drifting spiral waves as a mechanism for ventricular fibrillation. *Circulation* **94**, 1–48 (1996).
9. Pertsov, A. M., Davidenko, J. M., Salomonsz, R., Baxter, W. T. & Jalife, J. Spiral waves of excitation underlie reentrant activity in isolated cardiac muscle. *Circ. Res.* **72**, 631–650 (1993).
10. Winfree, A. T. *When Time breaks down* 40, 292 (Princeton Univ. Press, Princeton, NJ, 1987).
11. Wiggers, C. J. Studies of ventricular fibrillation produced by electric shock. II. Cinematographic and electrocardiographic observations of the natural process in the dog's heart. Its inhibition by potassium and the revival of coordinated beats by calcium. *Am. Heart J.* **5**, 351–365 (1930).
12. Witkowski, F. X. *et al.* A method for visualization of ventricular fibrillation: design of a cooled fiberoptically coupled image intensified CCD data acquisition system incorporating wavelet shrinkage based adaptive filtering. *Chaos* **8**, 94–102 (1998).
13. Chen, P.-S. *et al.* Mechanism of ventricular vulnerability to single premature stimuli in open-chest dogs. *Circ. Res.* **62**, 1191–1209 (1988).
14. Witkowski, F. X., Plonsey, R., Penkoske, P. A. & Kavanagh, K. M. Significance of inwardly directed transmembrane current in determination of local myocardial electrical activation during ventricular fibrillation. *Circ. Res.* **74**, 507–524 (1994).
15. Russ, J. C. *The Imaging Processing Handbook* 2nd edn 157–158 (CRC Press, New York, 1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>).

Acknowledgements. We acknowledge support from MRC grants (to F.X.W., P.A.P., L.J.L. and W.R.G.), the Alberta Heritage Foundation for Medical Research (F.X.W. and W.R.G.), the Office of Naval Research Physical Sciences Division (M.L.S. and W.L.D.), the NSWC ILIR program (M.L.S.), the Georgia Research Alliance (W.L.D.) and the National Science Foundation (A.T.W.).

Correspondence and requests for materials should be addressed to F.X.W. (e-mail: fwitkows@gr.u.ualberta.ca).

Synapse-specific control of synaptic efficacy at the terminals of a single neuron

Graeme W. Davis & Corey S. Goodman

Howard Hughes Medical Institute, Neurobiology Division, Department of Molecular and Cell Biology, LSA room 519, University of California, Berkeley, California 94720, USA

The regulation of synaptic efficacy is essential for the proper functioning of neural circuits. If synaptic gain is set too high or too low, cells are either activated inappropriately or remain silent. There is extra complexity because synapses are not static, but form, retract, expand, strengthen, and weaken throughout life. Homeostatic regulatory mechanisms that control synaptic efficacy presumably exist to ensure that neurons remain functional within a meaningful physiological range^{1–5}. One of the best defined systems for analysis of the mechanisms that regulate synaptic efficacy is the neuromuscular junction. It has been shown, in organisms ranging from insects to humans, that changes in synaptic efficacy are tightly coupled to changes in muscle size during development^{1,6–8}. It has been proposed that a signal from muscle to motor neuron maintains this coupling⁹. Here we show, by genetically manipulating muscle innervation, that there are two independent mechanisms by which muscle regulates synaptic efficacy at the terminals of single motor neurons. Increased muscle innervation results in a compensatory, target-specific decrease in presynaptic transmitter release, implying a retrograde regulation of presynaptic release. Decreased muscle innervation results in a compensatory increase in quantal size.

In each abdominal hemisegment of the *Drosophila* embryo and larva, roughly 40 motor neurons specifically synapse with 30

identified muscle fibres. For example, muscles 6 and 7 are both specifically innervated by motor neuron RP3 (Fig. 1a), muscle 12 is innervated by motor neuron RP5, and muscle 13 is innervated by motor neuron RP1 (Fig. 1b). A common exciter (motor neuron 6/7) innervates all four of these muscles¹⁰.

This stereotyped pattern of synaptic connections can be controlled by the differential expression of the cell adhesion molecule fasciclin II (refs 11–13). Motor neurons RP3 and 6/7 normally distribute their synaptic boutons fairly evenly between muscles 6 and 7 (Fig. 1a). When fasciclin II is overexpressed (using the Gal4-upstream activation sequence (UAS) system^{14,15}) on muscle 6 but not muscle 7, synapse formation and synaptic growth from motor neurons RP3 and 6/7 are strongly biased towards muscle 6, although the total bouton number of these motor neurons does not change (Fig. 1c)¹³. Ectopic synapses participate in muscle depolarization, as shown by both whole-cell and single-bouton recordings¹³. Because of both the biased growth of normal synapses and the formation of ectopic synapses, muscle 6 becomes hyperinnervated, receiving 91% more boutons than normal. Muscle 7 (the other target of motor neurons RP3 and 6/7), on the other hand, receives 40% fewer boutons than normal when fasciclin II is overexpressed on muscle 6 (Fig. 1c).

A similar situation is seen when fasciclin II is overexpressed on muscle 13 but not on neighbouring muscle 12. Muscle 13 becomes hyperinnervated, now receiving the majority of synapses from both motor neuron RP1 and motor neuron RP5 (162% increase; motor neuron RP5 normally synapses only with muscle 12), whereas muscle 12 receives 34% fewer boutons (Fig. 1d). Thus, by manipulating fasciclin II levels, we can create situations in which motor neurons RP3 and RP5 are each confronted with two targets with altered innervation. Two muscle targets are hyperinnervated (muscle 6, and muscle 13), whereas two other adjacent muscle targets are hypo-innervated (muscle 7 and muscle 12).

The physiological consequences of these changes in synaptic connectivity show two mechanisms by which muscle regulates synaptic efficacy. Intracellular recordings from hyperinnervated muscles 6 and 13 during nerve stimulation show that muscle depolarization is normal despite the dramatic increase in bouton number (Fig. 2). Quantal size is normal at these synapses, as estimated by studying the mean spontaneous miniature excitatory junctional current (MEJC). On average, there is a small increase in quantal content (the average number of presynaptic vesicles released per stimulus) at these muscles (27% increase at muscle 6 and 33% increase at muscle 13, as determined by dividing the average EJC amplitude by the average MEJC amplitude; Fig. 2). However, this modest increase in quantal content is far less than what one would expect, given a 91 or 162% increase in bouton number.

Two alternative mechanisms could account for normal synaptic output despite increased bouton number. Many of the boutons contacting these muscles could be silent. Alternatively, the probability of transmitter release at each presynaptic bouton could be downregulated. To distinguish between these alternatives, we quantified transmitter release at single synaptic boutons. Extracellular recordings from single boutons show that silent synaptic boutons are not present either at wild-type muscles 6 and 13 or at fasciclin-II-induced hyperinnervated muscles 6 and 13. These data concur with the observation that ectopic boutons cluster postsynaptic glutamate receptors (data not shown). Instead, there is a significant reduction in the probability of transmitter release at boutons contacting hyperinnervated muscles 6 and 13 compared with the probability of release during the innervation of wild-type muscles (Fig. 3a–c). The boutons producing both the normal innervation (by motor neurons RP3 and 6/7) and the ectopic innervation have a similar probability of transmitter release at hyperinnervated muscle 6, indicating that all inputs to a single muscle may be similarly regulated (Fig. 3a). However, as there are no data on the release properties of ectopic synapses at their normal targets, we cannot

conclude that the probability of release at ectopic synapses has been changed.

We also estimated the per-bouton quantal content by the method of failures for data recorded in low calcium (0.25 mM) at muscle 13 (Fig. 3b). Significantly, the observed percentage decrease in the per-bouton quantal content is enough to completely compensate for the increase in bouton number at muscle 13.

This muscle-to-motor-neuron signalling is also target-specific. Overexpression of fasciclin II on muscle 6 stabilizes more boutons

from motor neuron RP3 onto muscle 6 and fewer boutons from motor neuron RP3 onto muscle 7. When single boutons were sampled at muscle 7 in these experiments, these boutons had a normal probability of transmitter release (Fig. 3c). This demonstrates that the retrograde signal that downregulates presynaptic release at the boutons contacting hyperinnervated muscle 6 does not spread to the boutons from the same motor neuron that contact muscle 7.

To test whether the level of fasciclin II has a direct effect on the

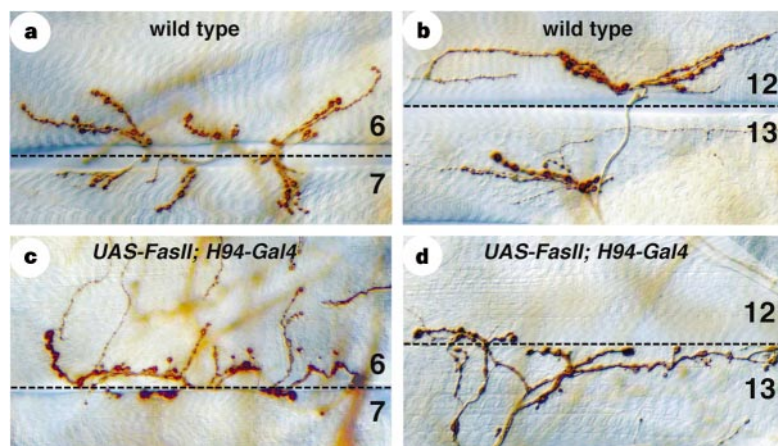


Figure 1 Differential expression of fasciclin II (FasII) directs synapse formation and growth. **a**, Synaptic boutons from motor neurons RP3 and 6/7 are reliably distributed 40% and 60%, respectively, on muscles 7 and 6 (ref. 13). The dotted line in this and all other panels delineates the separation between muscles. **b**, Motor neuron RP1 synapses with muscle 13, whereas motor neuron RP5 synapses with muscle 12. Motor neuron RP5 crosses over muscle 13 before selectively synapsing with muscle 12. **c**, Overexpression of FasII on muscle 6 in *UAS-FasII;H94-Gal4* larvae biases synapse formation and growth towards muscle 6 and away from muscle 7. FasII overexpression also stabilizes novel, ectopic synapses on muscle 6 but not muscle 7. **d**, Overexpression of FasII in muscle 13 by *H94-Gal4* stabilizes synaptic boutons from motor neuron RP5 that would normally contact muscle 12 on muscle 13.

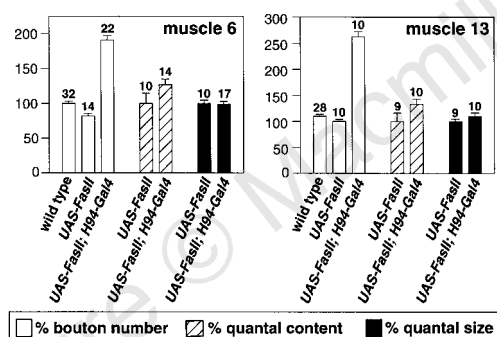


Figure 2 Physiological consequences of increased muscle innervation. Quantification of bouton number, quantal size and quantal content (normalized to wild type and expressed as a percentage) for the synapse on muscles 6 (left) and 13 (right) in wild-type, *UAS-FasII* and *UAS-FasII;H94-Gal4* larvae. There is a small, statistically significant, increase in quantal content that accompanies the large increase in bouton number at muscles 6 and 13 ($P < 0.03$, Student's *t*-test). There is no significant change in quantal size. All data are mean $\pm$ s.e.m.

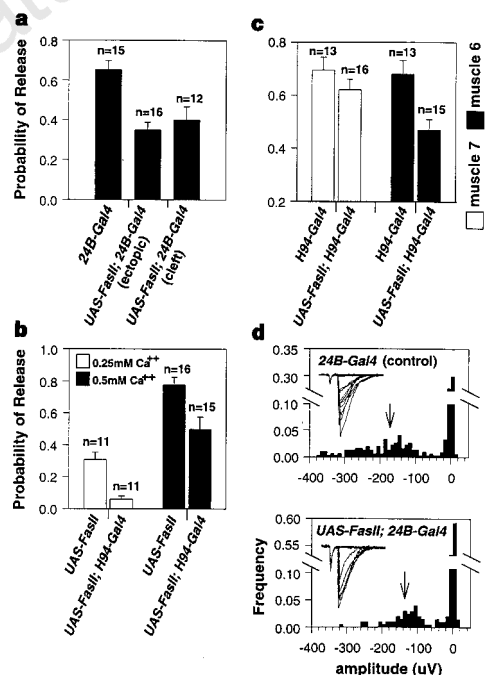


Figure 3 The per-bouton probability of transmitter release is decreased at superinnervated muscles. **a**, The probability of observing a release event (probability of release) was quantified for single boutons at muscles 6 in wild-type (*24B-Gal4*) and in *UAS-FasII;24B-Gal4* larvae. Boutons were sampled from the cleft between muscles 6 and 7 (cleft), which is normally innervated by motor neurons RP3 and 6/7. Ectopic boutons were also sampled at the dorsal edge of muscle 6, which normally does not receive synaptic input (ectopic). Transmitter release from single boutons contacting muscle 6 decreased in *UAS-FasII;24B-Gal4* larvae compared with in *24B-Gal4* larvae ($P < 0.005$; Student's *t*-test). There was no change in release probability when comparing expression of *UAS-FasII* (59.1 ± 1.2 , $n = 11$) and *UAS-FasII;MHC-Gal4* (62.5 ± 1.5 , $n = 11$; data not shown). **b**, The probability of release at single boutons contacting hyperinnervated muscle 13 (*UAS-FasII;H94-Gal4*) is significantly reduced compared with the release probability at wild-type muscle 13 (*UAS-FasII*; $P < 0.001$) in two different Ca^{2+} concentrations. **c**, Increased bouton number at muscle 6 in *UAS-FasII;H94-Gal4* larvae compared with *H94-Gal4* larvae results in a significantly decreased probability of release at the boutons from motor neuron RP3 on muscle 6 ($P < 0.01$). There was no change at boutons from the same motor neurons at muscle 7 (with decreased innervation). **d**, Sample amplitude histograms from boutons contacting wild-type muscle 6 (*24B-Gal4*; $n = 300$) and hyperinnervated muscle 6 (*UAS-FasII;24B-Gal4*; $n = 220$). Vertical arrows indicate the average spontaneous release event amplitude. Insets are sample traces from each genotype (30 traces superimposed).

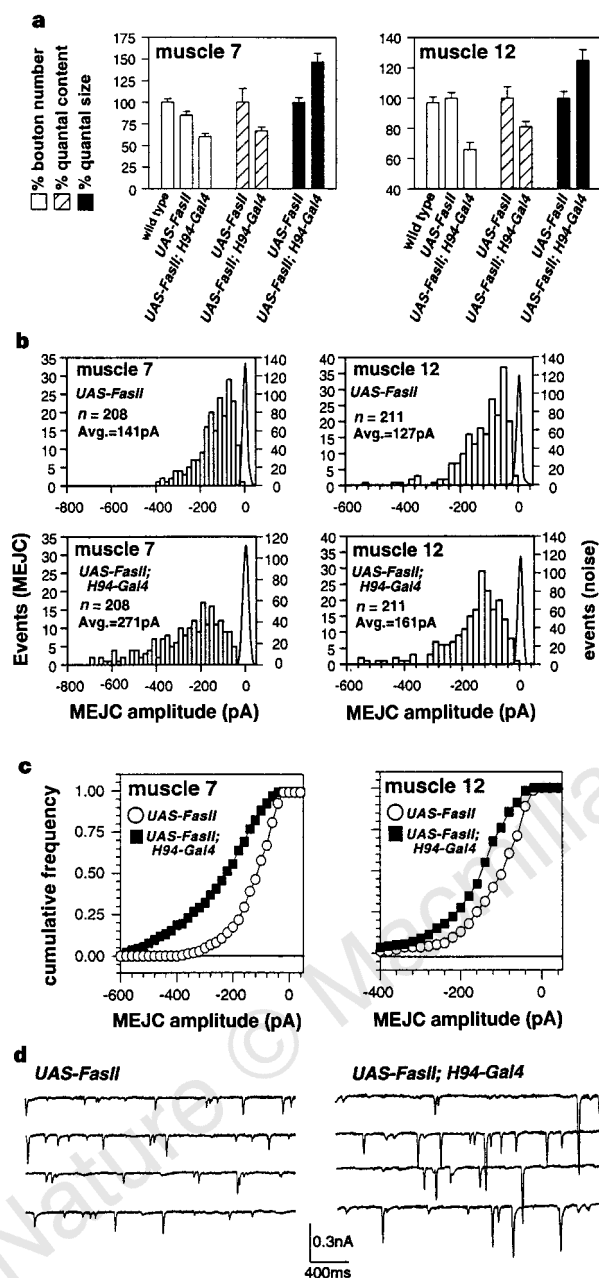


Figure 4 Quantal size is increased at muscles that receive reduced innervation. **a**, In *UAS-FasII;H94-Gal4* larvae there is a significant reduction in both bouton number ($P < 0.001$; Student's t -test; muscles 7 and 12) and quantal content ($P < 0.01$, muscles 7 and 12). There is a significant increase in quantal size ($P < 0.001$, muscle 7; $P < 0.005$, muscle 12). Data are normalized to wild type. **b**, Representative amplitude histograms for MEJC amplitudes recorded in muscle 7 or muscle 12, comparing wild-type (*UAS-FasII*) with hypo-innervated (*UAS-FasII;H94-Gal4*) junctions. Noise levels are indicated by solid lines around zero amplitude. **c**, Cumulative probability plots of data for the population of synapses in each genotype at muscle 7 and muscle 12. There was no change in quantal size at muscle 7 (comparing wild-type (*UAS-FasII*) and *UAS-FasII;H94-Gal4* larvae (7.6%, $P = 0.39$)). **d**, Sample traces from muscle 7 in wild-type larvae (*UAS-FasII*) and when innervation is decreased (*UAS-FasII;H94-Gal4*). There was no change in the intersegmental coupling between muscles 7, comparing *UAS-FasII* muscles ($49.9 \pm 5.1\%$ voltage attenuation, $n = 10$) to hypo-innervated muscles (*UAS-FasII;H94-Gal4*; $46.3 \pm 5.0\%$ attenuation, $n = 11$; data not shown). There was no change in the average resting potential, comparing *UAS-FasII* larvae (muscle 7, -65.2 ± 2.9 mV; muscle 12, -61.3 ± 2.0) and *UAS-FasII;H94-Gal4* larvae (muscle 7, -61.5 ± 2.9 mV; muscle 12, -58.6 ± 1.5 ; data not shown).

function of the synapse, we recorded from, and observed wild-type probability of release at, single boutons in a control genotype that overexpresses fasciclin II in all muscles throughout larval development (*UAS-FasII; MHC-Gal4*). This genotype expresses fasciclin II, with expression beginning at the first larval instar, which is after the critical period for synaptic rearrangement. The pattern of synaptic connections and the number of synaptic boutons are not altered.

We further examined the physiological effects of decreased innervation at neighbouring muscle targets contacted by motor neurons RP3 and RP5. Selective overexpression of fasciclin II on muscles 6 and 13 leads to a 35–40% reduction in innervation of muscles 7 and 12 by motor neurons RP3 and RP5 (Fig. 4a). At muscles 7 and 12, a significant reduction in quantal content accompanies this reduction in bouton number. These muscles respond to decreased innervation and decreased quantal content by upregulating quantal size (Fig. 4).

Several controls were used to allow comparison of quantal size between animals. *Drosophila* larval muscle fibres are small ($\sim 400 \mu\text{m}$ in length) and are isopotential¹⁶, aiding the comparison of quantal size between different animals. There was also no change in muscle size, resting potential, or intersegmental coupling between muscle fibres at hypo-innervated synapses (Fig. 4 legend). To study the effect of overexpression of fasciclin II itself on quantal size, we elevated fasciclin II levels on both muscle 6 and muscle 7, resulting in the formation of ectopic synapses on both muscles¹³. There was no significant effect on quantal size at muscle 7 in these larvae (Fig. 4c).

The increased quantal size seen at hypo-innervated muscle could represent a postsynaptic change in the sensitivity to transmitter (due to changes in receptor density or receptor sensitivity), or could be due to a change in presynaptic transmitter release (vesicle size or multiquantal release). Spontaneous multiquantal release would be expected to change the shape of the MEJC distribution and the quantal size. Amplitude histograms from single recordings (Fig. 4b) and from the population of synapses of each genotype (Fig. 4c) show that the quantal size is increased over the entire distribution, arguing against the possibility that increased quantal size is entirely due to spontaneous multiquantal release events. Amplitude distributions at hypo-innervated muscles were also compared with distributions at wild-type muscles by calculating the average distribution skewness and average variance/mean. Only at hypo-innervated muscle 7 was there a significant change in one of these parameters: a significant increase in the average variance/mean was seen (*UAS-FasII*, 42.6 ± 3.4 ; *UAS-FasII; H94-Gal4*, 65.1 ± 8.4), but skewness did not change (Kruskal-Wallis H -test). This increased variance could reflect the previous emergence of release events within the noise at hypo-innervated muscles 7, or the emergence of a class of larger multiquantal release events not obviously revealed in amplitude histograms. We have performed controls that show the following: first, the MEJCs are independent, non-interacting events (Fig. 5); second, that the random summation of MEJCs is rare (see Methods); and third, that two independent methods of analysis (EJC/MEJC and failure analysis at single boutons) support similar conclusions. Ultimately, we cannot distinguish between a change in postsynaptic sensitivity to transmitter and a change in the filling or size of presynaptic vesicles.

We have shown that two independent mechanisms regulate synaptic gain at the terminals of single motor neurons. A muscle-to-motor-neuron retrograde regulation of presynaptic release downregulates synaptic efficacy at motor neuron terminals that contact muscles that receive increase innervation. An increase in quantal size enhances synaptic efficacy at the terminals from the same motor neurons that contact muscle targets receiving reduced innervation. These regulatory mechanisms are different from the presumably cell-autonomous homeostatic mechanisms that have been previously seen at these synapses^{8,17,18}. Understanding how neurons monitor their excitability and regulate the efficacy of their

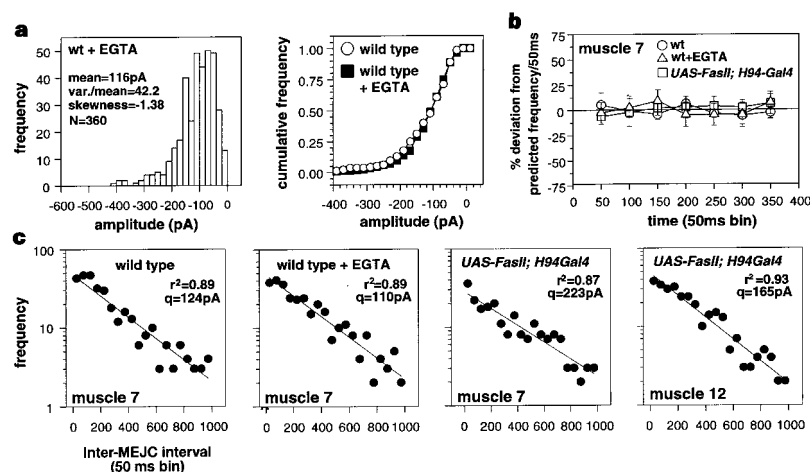


Figure 5 Spontaneous MEJCs are independent, non-interacting events. **a**, Representative amplitude histogram for data recorded in zero extracellular Ca^{2+} (0 mM Ca^{2+} , 4 mM EGTA) and cumulative probability distributions of the population of synapses recorded at 0 mM Ca^{2+} (closed squares) and 5 mM Ca^{2+} (open circles). There was no change in average skewness (0 mM Ca^{2+} = -1.44 ± 0.3 ; 5 mM Ca^{2+} = -1.20 ± 0.3 ; Kruskal-Wallis H -test) or in the average variance/mean (0 mM Ca^{2+} = 43.8 ± 3.8 ; 5 mM Ca^{2+} = 37.1 ± 4.1 ; Student's t -test, Mann-Whitney U -test). WT, wild type. **b**, The independence of MEJC events is not affected by extracellular calcium (5 mM Ca^{2+} (WT) versus 0 mM Ca^{2+} (WT + EGTA)) or by changes in quantal size (UAS-FasII;H94-Gal4) at muscles 7. **c**, Representative examples from four genotypes (muscle 7) show that MEJCs are independent events in time. The distances between sequential MEJC events were measured and binned at 50-ms intervals.

synaptic input may have relevance for a variety of clinical disorders. Disruption of these homeostatic mechanisms might adversely alter the function of neural circuits, resulting in seizure-like activity. Moreover, homeostatic mechanisms that normally ensure a balance of synaptic efficacy in the central nervous system might lead to various clinical syndromes of the injured central nervous system. For example, spinal cord injury would lead to the removal of major classes of descending inputs and, as a result, homeostatic mechanisms controlling synaptic efficacy might lead to the observed inappropriate strengthening of persisting inputs. □

Methods

Third-instar *Drosophila* larvae were selected at the wall-climbing stage and dissected as described⁸. Physiological preparations were fixed after recording and processed for immunostaining using described methods⁸. The serum antibody against synaptotagmin (provided by T. Littleton and H. Bellen) was used at a final dilution of 1:2000. The monoclonal antibody against fasciclin II (monoclonal antibody 1D4; G. Helt and C. S. G., unpublished observations) was used at a dilution of 1:5. Antibodies against *Drosophila* glutamate receptors were provided by Y. Kidokuro.

Genetics. Fasciclin II was expressed transgenically using the Gal4-UAS expression system^{14,15}. Transgenic expression of fasciclin II in subsets of muscle and nerve was achieved by crossing UAS-FasII homozygous flies (PEST⁺, transmembrane isoform) with H94-Gal4, MHC⁸²-Gal4, 24B-Gal4 or E105-Gal4 flies. H94-Gal4 flies express fasciclin II at high levels in muscles 4, 6 and 13 and in neurons in the intersegmental nerve and segmental nerve (D. Lin, personal communication). MHC⁸²-Gal4 drives expression of fasciclin II at high levels specifically in muscles, with expression beginning at the first larval instar (M. Winberg, personal communication). E105-Gal4 expresses fasciclin II in a subset of muscles and nerves, including in muscles 6 and 7 (D. Lin, personal communication). H94-Gal4, 24B-Gal4, and E105-Gal4 all drive expression of fasciclin II in muscle during the period of growth-cone exploration and synapse formation (embryonic stages 13–17). Overexpression of UAS-FasII by any of these three Gal4 lines specifically alters synapse formation at the muscles that overexpress Fas II. MHC⁸²-Gal4 drives UAS-FasII expression at high levels in muscle, but not until after synapse formation is complete, and it does not alter the pattern of synaptic connectivity¹³.

Physiology. For physiological recordings, larvae were dissected in ice-cold haemolymph-like saline (HL3) (ref. 19) as described¹⁸. Recordings were made using sharp microelectrodes (borosilicate glass, 1 mm optical density), with a resistance of 12–15 M Ω , filled with either 3 M KCl or 3 M potassium acetate. Data were digitized and recorded to disk using a Digidata 1200 analogue-to-digital board and PCLAMP software (Axon Instruments). Only recordings with resting membrane potentials of between -50 and -70 mV were selected for data acquisition. Stimulation of the segmental nerve was achieved by passing brief depolarizing pulses (of 100 μ s) with a MASTER-8 stimulus generator and stimulus isolation unit.

Single boutons were visualized with Nomarski optics ($\times 40$) and chosen for recording if they could be clearly isolated from the chain of boutons running along the surface of the muscle fibre. A large-bore patch electrode, with a diameter slightly smaller than that of the smaller type 1s boutons, was placed directly over the isolated bouton. The patch electrode was filled with 0.5 mM or 0.25 mM Ca^{2+} HL3 saline, the same saline used to bathe the preparation for each recording. The bouton activity was recorded using an Axopatch 2000 (Axon Instruments) in fast-clamp mode. Between 100 and 300 stimuli were recorded and analysed for each synaptic bouton.

Quantal analysis. *Drosophila* muscles are isopotential¹⁶. The amplitudes of spontaneous MEJCs were measured as described⁸. MEJCs with slow time courses, unambiguously originating from neighbouring segments, were excluded from analysis on the basis of the MEJC kinetics of rise time and decay. Background noise was measured by placing measurement cursors at a fixed separation equal to the MEJC rise time and by then recording background noise at a fixed position on each recorded trace. Failure analysis at single boutons was performed as described¹³, except that recordings were performed in 0.25 mM Ca^{2+} . Roughly 100 MEJs from each recording in a single genotype were combined to generate cumulative probability plots for the population of synapses within a single genotype. The skewness and mean/variance of MEJC amplitudes were calculated for each recording and then averaged for each genotype.

To support our use of quantal analysis at this synapse we used several controls. First, we demonstrated that MEJCs are independent, non-interacting events, despite their skewed amplitude distribution at the synapse¹⁶, and that this assertion holds true at 0 mM and 5 mM extracellular Ca^{2+} and when quantal size is increased at hypo-innervated muscle (Fig. 5). The occurrence of MEJC events was quantified within 50-ms bins of non-overlapping 300-ms segments of data. The percentage deviation from the predicted frequency per 50-ms bin, based on the MEJC rate, was determined for each recording. The average percentage deviation from the predicted frequency per 50 ms is plotted for each genotype tested (Fig. 5b).

Second, we estimated that the indistinguishable summation of two MEJCs comprises $<1.2\%$ of all spontaneous events at wild-type synapses, and $<1\%$ of events at hypo-innervated muscles where the quantal size is increased and the MEJC rate is decreased. The likelihood that two, independent MEJCs summate indistinguishably was estimated by T_r/T_i , where T_r is the average 10–90% MEJC rise time for a single recording and T_i is the average time interval between successive MEJCs derived from the MEJC rate for each recording. On average, $T_r/T_i = 1.2\%$ for wild-type muscles 7 in 0.5 mM Ca^{2+} , and 0.9% for hypo-innervated muscles 7 in 0.5 mM Ca^{2+} ; $T_r/T_i = 1.4\%$ for wild-type muscles 12 and 1.1% for hypo-innervated muscles 12.

Third, MEJC amplitude distributions recorded in zero calcium (0 Ca^{2+} , 4 mM EGTA; this abolishes evoked release) are very similar to those recorded in normal (5 mM) external Ca^{2+} (Fig. 5a). No significant difference was seen when comparing the cumulative probability distributions for the population of synapses at each calcium concentration, by comparing the average MEJC

variance/mean, or by comparing the average MEJC amplitude distribution skewness (Fig. 5a). This indicates that large spontaneous events are not due to the nonrandom coupling of multiple spontaneous release events by gated influx of calcium into the presynaptic terminal. However, we cannot discount the possibility that calcium fluxes from intracellular stores cause the nonrandom spontaneous release of multiple vesicles.

Received 10 September 1997; accepted 12 January 1998.

1. Purves, D. *Body and Brain: A Trophic Theory of Neural Connections* (Harvard Univ. Press, 1988).
2. Miller, K. D. Synaptic economics: competition and cooperation in synaptic plasticity. *Neuron* **17**, 371–374 (1996).
3. Abraham, W. C. & Bear, M. F. Metaplasticity: the plasticity of synaptic plasticity. *Trends Neurosci.* **19**, 126–131 (1996).
4. Bear, M. F. A synaptic basis for memory storage in the cerebral cortex. *Proc. Natl Acad. Sci. USA* **93**, 13453–13459 (1996).
5. Marder, E., Abbott, L. F., Turrigiano, G. G., Zheng, L. & Golowasch, J. Memory from the dynamics of intrinsic membrane currents. *Proc. Natl Acad. Sci. USA* **93**, 13481–13486 (1996).
6. Magrassi, L., Purves, D. & Lichtman, J. W. Visualization of neuromuscular junctions over periods of several months in living mice. *J. Neurosci.* **7**, 1215–1222 (1987).
7. Govind, C. K. & Pierce, J. Remodelling of multiterminal innervation by nerve terminal sprouting in an identifiable lobster motoneuron. *Science* **212**, 1522–1524 (1981).
8. Schuster, C. M., Davis, G. W., Fetter, R. F. & Goodman, C. S. Genetic dissection of structural and functional components of synaptic plasticity: Fasciclin II controls synaptic stabilization and growth. *Neuron* **17**, 641–654 (1996).
9. Frank, E. Matching of facilitation at the neuromuscular junction of the lobster: a possible case for influence of muscle on nerve. *J. Physiol.* **233**, 635–658 (1973).
10. Keshishian, H. *et al.* Cellular mechanisms governing synaptic development in *Drosophila melanogaster*. *J. Neurobiol.* **24**, 575–787 (1995).
11. Grenningloh, G., Rehm, E. J. & Goodman, C. S. Genetic analysis of growth cone guidance in *Drosophila*: Fasciclin II functions as a neuronal recognition molecule. *Cell* **67**, 45–57 (1991).
12. Lin, D. M. & Goodman, C. S. Ectopic and increased expression of Fasciclin II alters motoneuron growth cone guidance. *Neuron* **13**, 507–523 (1994).
13. Davis, G. W., Schuster, C. M. & Goodman, C. S. Genetic analysis of the molecular mechanisms controlling target selection: target derived Fasciclin II regulates the pattern of synapse formation. *Neuron* **19**, 561–573 (1997).
14. O'Kane, C. & Gehring, W. J. Detection *in situ* of genomic regulatory elements in *drosophila*. *Proc. Natl Acad. Sci. USA* **84**, 9123–9127 (1987).
15. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
16. Jan, L. Y. & Jan, Y. Y. Properties of the larval neuromuscular junction in *Drosophila melanogaster*. *J. Physiol. (Lond.)* **262**, 189–214 (1976).
17. Stewart, B. A., Schuster, C. M., Goodman, C. S. & Atwood, H. L. Homeostasis of synaptic transmission in *Drosophila* with genetically altered nerve terminal morphology. *J. Neurosci.* **16**, 3877–3886 (1996).
18. Schuster, C. M., Davis, G. W. & Goodman, C. S. Genetic dissection of structural and functional components of synaptic plasticity: Fasciclin II controls structural plasticity. *Neuron* **17**, 655–667 (1996).
19. Stewart, B. A., Atwood, H. L., Renger, J. J., Wang, J. & Wu, C. F. Improved stability of *Drosophila* larval neuromuscular preparations in haemolymph-like physiological solutions. *J. Comp. Physiol.* **175**, 175–191 (1994).

Acknowledgements. We thank R. S. Zucker for technical advice and encouragement; C. Schuster for advice during early stages of this project; D. Lin for Gal-4 lines; and K. Zito and A. DiAntonio for reading the manuscript. G.W.D. is a postdoctoral associate and C.S.G. is an Investigator with the HHMI. This work was supported by an NIH grant (to C.S.G.).

Correspondence and requests for materials should be addressed to G.W.D. (e-mail: gdavis@coreys.berkeley.edu).

Dendritic cells acquire antigen from apoptotic cells and induce class I-restricted CTLs

Matthew L. Albert, Birthe Sauter & Nina Bhardwaj

Laboratory of Cellular Physiology and Immunology, The Rockefeller University, New York, New York 10021, USA

CD8⁺ cytotoxic T lymphocytes (CTLs) mediate resistance to infectious agents and tumours. Classically, CTLs recognize antigens that are localized in the cytoplasm of target cells, processed and presented as peptide complexes with class I molecules of the major histocompatibility complex (MHC)¹. However, there is evidence for an exogenous pathway whereby antigens that are not expected to gain access to the cytoplasm are presented on MHC class I molecules^{2–6}. The most dramatic example is the *in vivo* phenomenon of cross-priming⁷: antigens from donor cells are acquired by bone-marrow-derived host antigen-presenting cells (APCs) and presented on MHC class I molecules. Two unanswered questions concern the identity of this bone-marrow-derived cell and how such antigens are acquired. Here

we show that human dendritic cells, but not macrophages, efficiently present antigen derived from apoptotic cells, stimulating class I-restricted CD8⁺ CTLs. Our findings suggest a mechanism by which potent APCs acquire antigens from tumours, transplants, infected cells, or even self-tissue, for stimulation or tolerization of CTLs.

Influenza A virus establishes a non-toxic infection in human dendritic cells (DCs)^{8,9}. Once infected, these DCs are capable of eliciting virus-specific, recall CTL responses within 7 days. The CTLs generated are class I-restricted and kill virus-infected monocytes and peptide-pulsed target cells^{8,10}. Compared with DCs, influenza-infected monocytes are poor stimulators of CTL responses⁸ and undergo apoptosis^{11,12}. We have exploited these observations to investigate the role of apoptosis in the generation of antigen that could be acquired by uninfected DCs.

DCs were prepared from peripheral blood precursors of HLA-A2.1⁺ donors^{13,14}. Uninfected DCs and syngeneic T cells were cocultured with influenza-infected syngeneic (Fig. 1a) or allogeneic (Fig. 1b) monocytes for 7 days. Influenza-specific CTLs were generated in these cocultures, suggesting that the DCs acquired antigen from the monocytes. The DCs and not the infected monocytes functioned as the antigen-presenting cell (APC), because the latter failed to stimulate CTLs in the absence of DCs, even at a stimulator:responder ratio of 1:2 (Fig. 1a, dashed lines). The CTL responses generated in these cocultures were as potent as those induced by influenza-infected DCs, and as few as 5×10^3 infected monocytes could charge uninfected DCs for the induction of robust CTLs (Fig. 1b). To confirm that antigen is transferred to and expressed directly by DC MHC class I products, we cocultured uninfected DCs and influenza-infected monocytes, and used the DCs as targets for influenza-specific CTLs (Fig. 1c). DCs cocultured with infected monocytes were recognized as efficiently as virus-infected DCs. Collectively, the results in Fig. 1 suggest that influenza antigen from infected monocytes gained access to MHC class I of the DC, that is, antigen from HLA-mismatched monocytes were cross-priming T cells through the DC.

It was important to exclude live virus as the agent responsible for the transfer of antigen to DCs. Influenza-infected monocytes produce only low levels of infectious virus before they undergo apoptosis^{9,11}. Infection of 5×10^4 monocytes with a multiplicity of infection (MOI) of 2.0 would be expected to yield up to 1×10^3 infectious virions after a 24 h culture period. To prevent infection of DCs by monocyte-produced virus, experiments were done in human serum, which contains blocking anti-haemagglutinin antibodies. Indeed, the addition of influenza virus (10^4 infectious virions) directly into cultures containing uninfected DCs and T cells did not yield a CTL response (data not shown). Furthermore, <1 haemagglutination units (HAU) ml⁻¹ was detected in the medium at 12 h, 24 h and 7 days of coculture.

The CTL responses were also not due to free peptide released by the dying monocytes, thereby charging class I molecules on DCs. Media from wells containing the infected monocytes were collected after an overnight culture and transferred to fresh wells containing T cells and uninfected DCs. Virus-specific CTLs were generated after a 7 day culture period (Fig. 2a, transfer). This CTL activity was abrogated if the medium was passed through a filter (0.45 µm pore size) before being added to T-cell–DC cultures (Fig. 2a, filter), suggesting that the antigenic material was neither live virus nor free peptide (as both would have passed through the filter). This was confirmed by sedimentation experiments (Fig. 2b). Media from wells containing infected monocytes were removed and spun at 250g. The antigenic material could be localized in the pellet, but not in the supernatant fraction. Notably, the pelleted material (which lacked detectable haemagglutinating activity) fully accounted for the CTL activity generated in direct transfer. These results are most consistent with the source of antigen being material derived from apoptotic cells.

The role of apoptosis in the transfer of antigen to the uninfected DC was therefore assessed. We first extended prior work¹¹, showing that influenza-infected monocytes undergo apoptosis as detected by annexin V binding, an early marker for programmed cell death (Fig. 3a, middle panel). In contrast, heat-inactivated (HI) influenza virus, which reliably reduces viral replication, failed to induce apoptosis in monocytes (Fig. 3a, bottom panel). However, DCs that are infected with HI influenza still stimulate potent influenza-specific CTL responses¹⁰. We employed this replication-deficient virus to probe the requirement for apoptosis in cross-priming. When media from wells containing the monocytes infected with HI influenza were collected after an overnight culture and transferred to fresh wells containing T cells and uninfected DCs, no influenza-specific CTLs were generated (Fig. 3b, 10 h transfer, HI Flu). In contrast, DCs exposed to media derived from live virus-infected cultures, which contained apoptotic cells, did elicit a CTL response (Fig. 3b, 10 h transfer, Live Flu). However, if the monocytes infected with HI influenza were allowed to undergo apoptosis spontaneously, as occurs during 7 days of culture¹⁵, antigenic material was generated and virus-specific CTLs were induced (Fig. 3b, coculture, HI Flu). This data also argues against the possibility that live virus within the apoptotic cell is responsible for the transfer of antigenic material to the DCs.

To establish that apoptosis is the critical trigger for cross-priming, we used Z-VAD-CHO, an irreversible peptide inhibitor of caspase

activity. Apoptosis of influenza-infected monocytes was blocked (determined by TUNEL), without affecting the expression of viral proteins (data not shown). Influenza-infected monocytes were cultured overnight in the presence of varying concentrations of Z-VAD. Media from these wells were then added to fresh wells containing T cells and uninfected DCs. At concentrations of Z-VAD which inhibited apoptosis of the infected monocytes, antigenic material was not transferred to the uninfected DCs, and virus-specific CTLs were not generated (Fig. 3c). We next compared apoptotic death to necrotic death for the generation of antigenic material. Influenza-infected 293 cells express influenza proteins after 10 h of infection but, unlike monocytes, do not undergo apoptosis. Apoptosis or necrosis was induced by ultraviolet B irradiation and hypotonic shock, respectively, and the cells were then cocultured with uninfected DCs and T cells for 7 days. CTL activity was measured using matrix peptide-pulsed T2 targets. Only the apoptotic cells and not the necrotic cells were capable of charging DCs with antigen (Fig. 3d). Apoptotic versus necrotic influenza-infected monocytes gave similar results (data not shown).

It was important to investigate the cellular requirements for generating CTLs through this pathway. When CD4⁺ and CD8⁺ subpopulations were purified at the end of the 7 day culture period, CTL activity was detected only in the CD8⁺ fraction (data not shown). We next compared DCs to macrophages as potential mediators of this exogenous pathway. Uninfected HLA-A2.1⁺

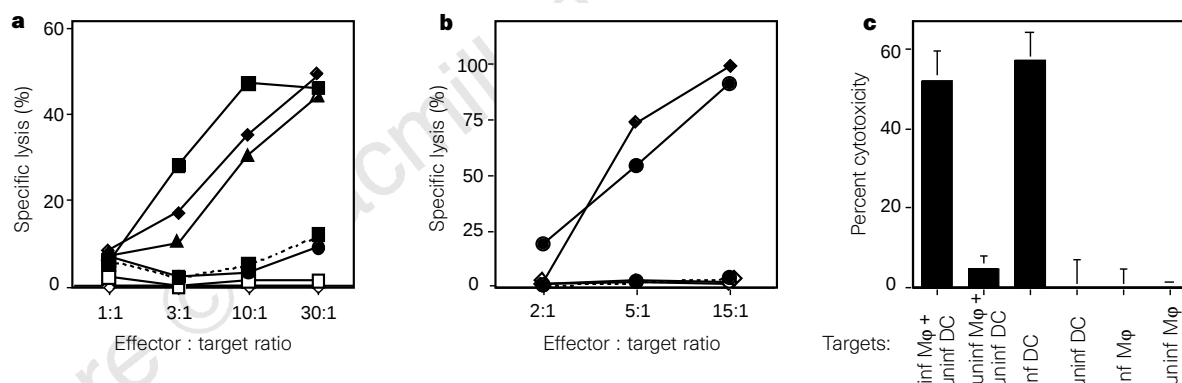


Figure 1 Dendritic cells acquire antigen from influenza-infected cells and induce class I-restricted CTLs. **a**, **b**, 8 to 10 h after infection with influenza virus, varying doses of syngeneic HLA-A2.1⁺ monocytes (**a**) or allogeneic HLA-A2.1⁺ monocytes (**b**) were added to cocultures of DCs and T cells. On day 7, cytolytic activity was tested using syngeneic influenza-infected macrophages (**a**) or T2 cells (a Tap^{-/-}, HLA-A2.1⁺ cell line) pulsed with the immunodominant influenza matrix peptide²⁶ (**b**) as targets. Symbols: filled squares plus solid line 5 × 10⁵ infected monocytes (inf Mφ) + uninfected dendritic cells (uninf DCs); open squares 5 × 10⁵ uninf Mφ + uninf DCs; filled squares plus dotted line, 5 × 10⁵ inf Mφ, no DCs; filled triangles,

5 × 10⁴ inf Mφ + uninf DCs; filled circles plus solid line, 5 × 10³ inf Mφ + uninf DCs; open circles, 5 × 10³ uninf Mφ + uninf DCs; filled circles plus dotted line, 5 × 10³ inf Mφ, no DCs; filled diamonds, inf DCs; open diamonds, uninf DCs. **c**, Uninfected syngeneic DCs are cocultured with influenza-infected or uninfected allogeneic monocytes for 2 days before being used as targets for CTLs. Control targets included influenza-infected syngeneic DCs and influenza-infected allogeneic monocytes. Effector:target ratio = 45:1. Results are representative of 8 experiments and the values shown represent the mean from triplicate wells.

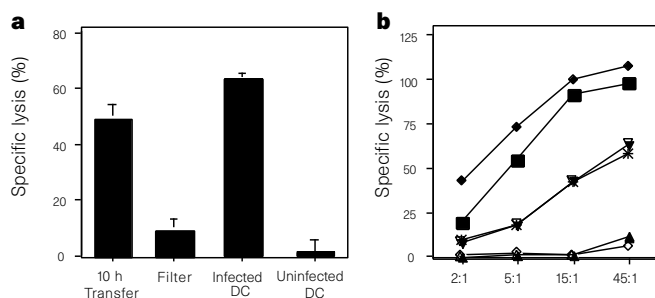


Figure 2 Antigen transfer is not due to live influenza virus or free peptide. **a**, 5 × 10⁴ infected allogeneic HLA-A2.1⁺ monocytes were cultured for 10 h, after which the media from these wells was collected. Media was either directly added to fresh wells containing HLA-A2.1⁺ T-cells and DCs (10 h transfer) or first passed through a 0.45 μm filter. Effector:target ratio = 30:1. **b**, 5 × 10³ infected allogeneic monocytes were cultured for 10 h, after which T cells and DCs were added to the wells (squares). Alternatively, media from wells containing infected monocytes was removed and transferred to fresh wells containing T cells and DCs (asterisks). Other cultures were established in which the medium was first spun at 250g in a GH-3.8 Beckman rotor for 10 min. The resulting supernatant fraction (triangles) versus the pellet (inverted triangles) was then added to DC-T cell cocultures. Controls included T cells cultured with infected (filled diamonds) and uninfected (open diamonds) DCs. After 7 days, cytolytic activities were determined on T2 cells pulsed with matrix peptide. Results are representative of 3 experiments.

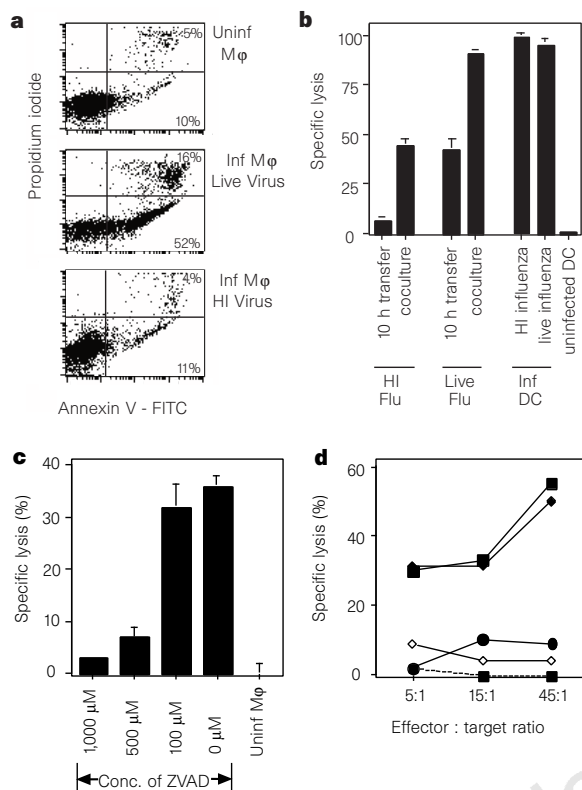


Figure 3 Apoptosis is required for delivery of antigen to DCs. **a**, Monocytes were infected with live or heat-inactivated influenza virus, cultured for 10 h and stained with annexin V – FITC (Kayima Biomedical Company, Seattle, WA) and propidium iodide. Early apoptotic cells are defined by the annexin V⁺, propidium iodide⁺ population. **b**, Allogeneic HLA-A2.1⁺ monocytes were infected with live or heat-inactivated virus. After 10 h, media from wells containing infected monocytes was removed and added to fresh wells containing T cells and uninfected DCs (10 h transfer). Alternatively, the infected monocytes were directly cocultured with the T-cells and DCs (coculture). Heat-inactivated influenza and live influenza-infected DCs served as the positive controls for the experiment. After 7 days CTL activity was measured on matrix peptide-pulsed targets. **c**, Apoptosis of influenza-infected monocytes was inhibited using Z-VAD-CHO (Kayima Biomedical Company). 1×10^4 infected monocytes were exposed to varying doses of Z-VAD for 1 h, washed and cultured for 10 h. Media from these wells was then transferred to fresh wells containing 2×10^5 T cells and 6.67×10^3 DCs. Cytolytic activity was determined as in **b**. **d**, 293 cells, a human kidney epithelial cell line, were infected with influenza virus, and cultured for 10 h. Apoptosis was induced by UVB irradiation for 2 min²⁸. Necrosis was achieved by incubating the 293 cells in a hypotonic solution for 30 min at 37°C, after which all the cells incorporated trypan blue. Apoptotic or necrotic 293 cells were cocultured with T cells and uninfected DCs for 7 days. Cytolytic activity was determined as in **b**. Uninfected 293 cells failed to induce CTL activity (data not shown). Symbols: squares plus solid line, apoptotic inf 293 cells + uninf DCs; circles, necrotic inf 293 cells + uninf DCs; squares plus dotted line, apoptotic inf 293 cells, no DCs; filled diamonds, inf DCs; open diamonds, uninf DCs. The results are representative of 9 experiments.

macrophages or DCs and syngeneic T cells were cocultured with infected HLA-A2.1⁺ monocytes for 7 days, after which cytolytic activity was measured. DCs but not macrophages were capable of stimulating influenza-specific CTLs (Fig. 4). Additionally, as increasing doses of syngeneic uninfected macrophages were introduced into cocultures containing uninfected DCs, T cells and infected allogeneic monocytes, the CTL activity was abrogated. Presumably, the macrophages act to sequester antigen from the DCs by efficiently engulfing the apoptotic cells and degrading the antigen.

To demonstrate that antigen was being processed intracellularly, DCs were pretreated with inhibitors of antigen presentation and

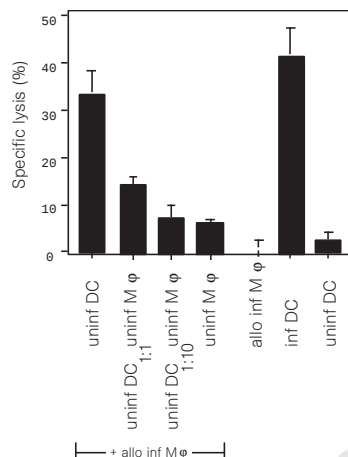


Figure 4 Dendritic cells and not macrophages are capable of cross-presenting apoptotic antigenic material. Uninfected HLA-A2.1⁺ monocytes or DCs and mixtures of both were used as APCs, and cocultured with bulk syngeneic T cells and influenza-infected HLA-A2.1⁺ DCs and infected HLA-2.1⁺ monocytes served as positive and negative controls, respectively. Cytolytic responses were measured after 7 days. Effector:target ratio = 30:1.

used as targets for influenza-specific CTLs. Both ammonium chloride and Brefeldin A completely inhibited the DC's ability to present antigenic material derived from apoptotic monocytes (data not shown). Lactacystin, an irreversible inhibitor of the 26S proteasome¹⁶, only partially blocked antigen presentation by the uninfected DCs (data not shown), suggesting that both classical and non-classical class I pathways are utilized for the presentation of exogenous antigen derived from apoptotic cells. We next documented the uptake of apoptotic cells by immunofluorescence and electron microscopy. Ten to twenty percent of the DCs contained fragmented or intact cell bodies in Fig. 5a. DCs were identified by the DC-restricted marker, p55, and apoptotic material was identified by the intense DAPI staining of pyknotic nuclei. This was confirmed by electron microscopy (Fig. 5b). About 20–25% of the DCs (identified by expression of CD83, data not shown) had vesicles containing apoptotic corpses with apparently intact plasma membranes. A recent report indicates that DCs associate with apoptotic cells through the vitronectin receptor $\alpha v \beta 3$, but fail to associate with opsonized particles and necrotic cells¹⁷. Further studies will be required to determine if DCs are actually phagocytosing and presenting apoptotic cells through this receptor.

Previous studies have shown that murine DCs have the capacity to present soluble antigens through an exogenous pathway, leading to the induction of MHC class I-restricted CTLs^{5,18–20}. Following intravenous injection of allogeneic cells into rats, interdigitating DCs within lymph nodes were found to contain whole cells and cell fragments²¹. Here we present a physiologically relevant system which demonstrates that human DCs can acquire relevant antigens and stimulate MHC class I-restricted CTLs by phagocytosing apoptotic cells. Our studies provide the first evidence that apoptosis (but not necrosis) is required for the generation and packaging of immunogenic material for delivery to the DC. We believe this pathway accounts for the *in vivo* phenomenon of 'cross-priming', whereby antigens derived from tumour cells²² or transplants²³ are presented by host APCs. Tolerance to tissue-restricted self antigens^{24,25} may also depend upon apoptotic cell death (as occurs during development and normal cell turnover) followed by antigen presentation by DCs. Possibly, heat shock proteins within the apoptotic cells direct antigen into the class I MHC presentation pathway⁴. As a consequence of apoptosis, antigens within cells that lack costimulatory function for T cells can gain access to the potent DC system, thereby eliciting stimulatory or tolerogenic responses. This apoptosis-dependent pathway has the potential to be therapeutically manipulated to induce CTL responses *in vivo* to a variety of antigens including tumour and microbial antigens, and possibly to modulate autoimmune responses. □

Methods

Generation of mononuclear subsets. Peripheral blood mononuclear cells

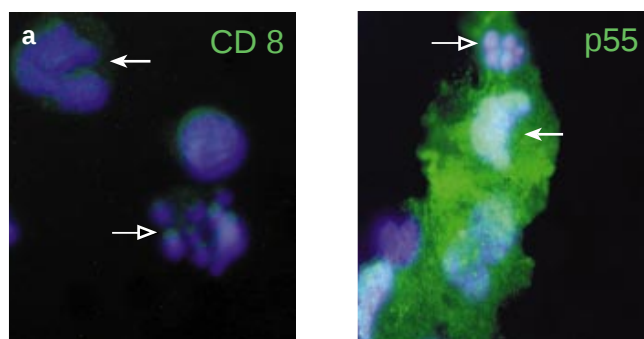
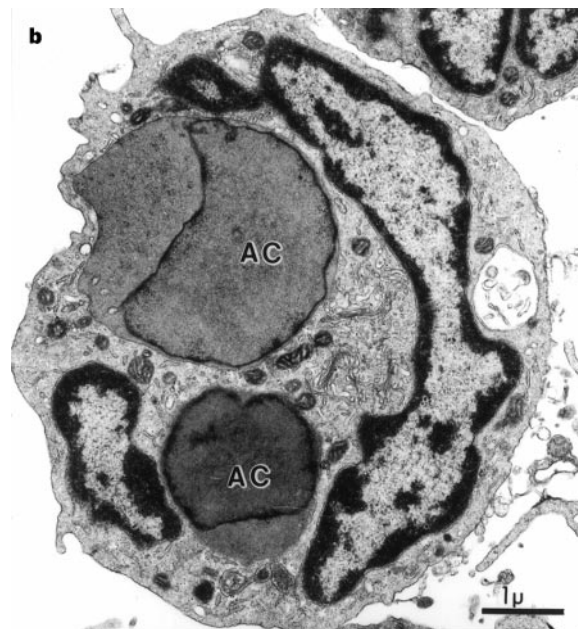


Figure 5 Dendritic cells engulf apoptotic monocytes. Influenza-infected monocytes and uninfected DCs were cocultured for up to 10 h. **a**, Cells were fixed with acetone and stained with anti-CD8 (isotype control) or anti-p55 followed by goat anti-mouse-FITC and incubation with DAPI (Sigma). The nuclei of the DCs (closed arrows) are lobulated and euchromatic as compared with the pyknotic, fragmented nucleus of the apoptotic cells (open arrows). DAPI⁺ material from an apoptotic cell appears to be within the cytoplasm of a p55⁺ DC. **b**, Electron microscopy revealed apoptotic material and apoptotic cells (AC) within the cytoplasm of glutaraldehyde-fixed DCs²⁹.



(PBMCs), T cells, and macrophages were prepared as previously described⁸. Dendritic cells were prepared from T-cell depleted PBMCs by culturing cells for 7 days in the presence of granulocyte and macrophage colony-stimulating factor (GM-CSF) and interleukin-4 (IL-4), followed by 4 days in monocyte-conditioned medium^{13,14}.

Induction and detection of apoptosis. Monocytes were infected with influenza virus in serum-free RPMI. Cell death was assayed using the Early Apoptosis Detection kit (Kayima Biomedical). Briefly, cells are stained with Annexin V-FITC (Ann V) and propidium iodide (PI). Early apoptosis is defined by Ann V⁺/PI⁻ staining as determined by FACSscan (Becton Dickinson). Cells from the 293 cell line were triggered to undergo apoptosis using a 60UVB lamp (Derma Control Inc.), calibrated to provide 2 mJ cm⁻² s⁻¹.

Coculture of DCs with apoptotic cells. Monocytes from HLA-A2.1⁻ donors were infected with live or heat-inactivated influenza virus. Live influenza virus (Spafas Inc.) was added at a final concentration of 250 HAU ml⁻¹ (MOI of 0.5) for 1 h at 37 °C (ref. 8). Virus was heat-inactivated by treatment for 30 min at 56 °C before use¹⁰. After washing, cells were added to 24-well plates in varying doses. After 1 h, contaminating non-adherent cells were removed and fresh media was added. Following a 10 h incubation at 37 °C, 3.3 × 10³ uninfected DCs and 1 × 10⁶ T cells were added to the wells.

Assay for virus-specific CTLs. After 7 days of culture, T cells were assayed for cytolytic activity using a conventional Na⁵¹CrO₄ release assay⁸. The targets were either influenza-infected syngeneic monocytes or T2 cells (a TAP^{-/-}, HLA-A2.1⁺, class II⁻ cell line) pulsed for 1 h with 1 μM of the immunodominant influenza matrix peptide, GILGFVFTL^{26,27}. Specific lysis was determined by subtracting the per cent killing of uninfected monocytes or unpulsed T2 cells. Influenza-infected DCs served as controls in all experiments in order to measure the donor's CTL responsiveness to influenza. Responses varied as a function of the individual's prior exposure to influenza. Background lysis ranged from 0–5% for the uninfected monocytes and 0–20% for the unpulsed T2 cells. Additionally, syngeneic DCs were used as targets where noted.

Received 3 October; accepted 1 December 1997.

- Heemels, M. T. & Ploegh, H. Generation, translocation, and presentation of MHC class I-restricted peptides. *Ann. Rev. Biochem.* **64**, 463–491 (1995).
- Watts, C. Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol.* **15**, 821–850 (1997).
- Kovacsics-Bankowski, M. & Rock, K. L. A phagosome-to-cytosol pathway for exogenous antigens presented on MHC class I molecules. *Science* **267**, 243–246 (1995).
- Suto, R. & Srivastava, P. K. A mechanism for the specific immunogenicity of heat shock-chaperoned peptides. *Science* **269**, 1585–1588 (1995).
- Norbury, C. C., Chambers, B. J., Prescott, A. R., Ljunggren, H.-G. & Watts, C. Constitutive macropinocytosis allows TAP-dependent presentation of exogenous antigen on class I MHC molecules by bone marrow derived dendritic cells. *Eur. J. Immunol.* **27**, 280–288 (1997).
- Pfeifer, J. D. et al. Phagocytic processing of bacterial antigens for class I MHC presentation to T cells. *Nature* **361**, 359–363 (1993).

- Bevan, M. J. Cross-priming for a secondary cytotoxic response to minor H antigens with H-2 congenic cells which do not cross-react in the cytotoxic assay. *J. Exp. Med.* **143**, 1283–1288 (1976).
- Bhardwaj, N. et al. Influenza virus-infected dendritic cells stimulate strong proliferative and cytolytic responses from human CD8⁺ T cells. *J. Clin. Invest.* **94**, 797–807 (1994).
- Bender, A. et al. The distinctive features of influenza virus infection of dendritic cells. *Immunobiology* (in the press).
- Bender, A., Bui, L. K., Feldman, M. A. V., Larsson, M. & Bhardwaj, N. Inactivated influenza virus, when presented on dendritic cells, elicits human CD8⁺ cytolytic T cell responses. *J. Exp. Med.* **182**, 1663–1671 (1995).
- Fesq, H., Bacher, M., Nain, M. & Gema, D. Programmed cell death (apoptosis) in human monocytes infected by influenza virus. *Immunobiology* **190**, 175–182 (1994).
- Hofmann, P. et al. Susceptibility of mononuclear phagocytes to influenza A virus infection and possible role in the antiviral response. *J. Leuk. Biol.* **612**, 408–414 (1997).
- Bender, A., Sapp, M., Schuler, G., Steinman, R. M. & Bhardwaj, N. Improved methods for the generation of dendritic cells from nonproliferating progenitors in human blood. *J. Immunol. Methods* **196**, 121–135 (1996).
- Romani, N. et al. Generation of mature dendritic cells from human blood: an improved method with special regard to clinical applicability. *J. Immunol. Methods* **196**, 137–151 (1996).
- Badley, A. D. et al. Macrophage-dependent apoptosis of CD4⁺ T lymphocytes from HIV-infected individuals is mediated by FasL and tumor necrosis factor. *J. Exp. Med.* **185**, 55–64 (1997).
- Fenteany, G. et al. Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science* **268**, 726–731 (1995).
- Rubartelli, A., Poggi, A. & Zocchi, M. R. The selective engulfment of apoptotic bodies by dendritic cells is mediated by the αvβ3 integrin and requires intracellular and extracellular calcium. *Eur. J. Immunol.* **27**, 1893–1900 (1997).
- Shen, Z., Reznikoff, G., Droanoff, G. & Rock, K. L. Cloned dendritic cells can present exogenous antigens on both MHC class I and class II molecules. *J. Immunol.* **158**, 2723–2730 (1997).
- Paglia, P., Chiodoni, C., Rodolfo, M. & Colombo, M. P. Murine dendritic cells loaded *in vitro* with soluble protein prime CTL against tumor antigen *in vivo*. *J. Exp. Med.* **183**, 317–322 (1996).
- Bachmann, M. F. et al. Dendritic cells process exogenous viral proteins and virus-like particles for class I presentation to CD8⁺ cytotoxic T lymphocytes. *Eur. J. Immunol.* **26**, 1–7 (1996).
- Fossum, S. & Rolstad, B. The roles of interdigitating cells and natural killer cells in the rapid rejection of allogeneic lymphocytes. *Eur. J. Immunol.* **16**, 440–450 (1986).
- Huang, A. Y. C. et al. Role of bone marrow-derived cells in presenting MHC class I-restricted tumor antigens. *Science* **264**, 961–965 (1994).
- Bevan, M. J. Priming for a cytotoxic response to minor histocompatibility antigens: antigen specificity and failure to demonstrate a carrier effect. *J. Immunol.* **118**, 1370–1374 (1977).
- Kurts, C. et al. Constitutive class I-restricted exogenous presentation of self antigens *in vivo*. *J. Exp. Med.* **184**, 923–930 (1996).
- Kurts, C., Kosaka, H., Carbone, F. R., Miller, J. F. A. P. & Heath, W. R. Class I-restricted cross-presentation of exogenous self antigens leads to deletion of autoreactive CD8⁺ T cells. *J. Exp. Med.* **186**, 239–245 (1997).
- Gotch, F., McMichael, A. & Rothbard, J. Recognition of influenza A matrix protein by HLA-A2-restricted cytotoxic T lymphocytes. *J. Exp. Med.* **168**, 2045–2057 (1988).
- Gotch, F., Rothbard, J., Howland, K., Townsend, A. & McMichael, A. Cytotoxic T lymphocytes recognize a fragment of influenza virus matrix protein in association with HLA-A2. *Nature* **326**, 881 (1987).
- Casciola-Rosen, L., Anhalt, G. & Rosen, A. Autoantigens targeted in systemic lupus erythematosus are clustered in two populations of surface structures on apoptotic keratinocytes. *J. Exp. Med.* **179**, 1317–1330 (1994).
- De Panfilis, G., Manara, G. C. & Ferrari, C. Hair cell leukemia cells express CD1a antigen. *J. Invest. Dermatol.* **87**, 510–514 (1986).

Acknowledgements. We thank R. Steinman, R. Darnell, Y. Choi and T. Sakmar for advice and critical review of the manuscript; G. Kaplan and W. Hellman for assistance with the EM; and B. van Steensel for assistance with the IF.

Correspondence and requests for materials should be addressed to N.B. (e-mail: bhardwn@rockvax.rockefeller.edu).

Role for interleukin-3 in mast-cell and basophil development and in immunity to parasites

Chris S. Lantz*, Jurg Boesiger*, Chang Ho Song*, Nicolas Mach†, Takahiko Kobayashi‡, Richard C. Mulligan§, Yukifumi Nawa‡, Glenn Dranoff† & Stephen J. Galli*

* Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, USA

† Department of Medicine, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Department of Parasitology, Miyazaki Medical College, Kiyotake, Miyazaki 889-16, Japan

§ Howard Hughes Medical Institute, Children's Hospital, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

The cytokine interleukin-3 (IL-3), which can be derived from T cells and other sources, is a potentially important link between the immune and haematopoietic systems¹. IL-3 may be particularly critical for the development, survival and function of tissue mast cells¹⁻⁶ and blood basophils^{7,8}, which are thought to be important effector cells in immunity to parasites and other immunological responses, such as allergic reactions⁹. Here we show, using IL-3 deficient mice¹⁰, that IL-3 is not essential for the generation of mast cells or basophils under physiological conditions, but that it

does contribute to increased numbers of tissue mast cells, enhanced basophil production, and immunity in mice infected with the nematode *Strongyloides venezuelensis*. Parasite expulsion and mast-cell development are impaired even more severely in IL-3-deficient mice that also show a marked reduction in signalling by c-kit. These findings establish a role for IL-3 in immunity to parasites and indicate that one of the functions of IL-3 in host defence against infection is to expand populations of haematopoietic effector cells.

Mice lacking IL-3 (IL-3^{-/-} mice) were produced using gene targeting in embryonic stem cells¹⁰. IL-3^{-/-} mice are healthy and fertile and, like mice that carry an inactivating mutation in the α -chain of the heterodimeric IL-3 receptor¹¹ or that lack both IL-3 and the common β -subunit of the receptors for IL-3, IL-5 and granulocyte-macrophage colony-stimulating factor¹², show no detectable abnormalities in multiple aspects of haematopoiesis *in vitro* or *in vivo*¹⁰. However, the c-kit ligand, stem cell factor (SCF)¹³, induced fewer mast cells to develop from bone-marrow cells of IL-3^{-/-} mice than were induced from bone-marrow cells of wild-type mice (Fig. 1a, b). In contrast, markedly higher numbers of mast cells developed when bone-marrow cells from either IL-3^{-/-} or wild-type mice were maintained in exogenous SCF plus IL-3 (Fig. 1b).

The findings shown in Fig. 1a, b concur with previous work indicating that exogenous IL-3 can augment SCF-dependent mast-cell development *in vitro*^{6,13,14}. Although numbers of mast-cell progenitors^{6,13-15} may differ in IL-3^{-/-} and IL-3^{+/+} mice *in vivo*, our data show that similar numbers of mast cells can be generated when bone-marrow cells from either IL-3^{-/-} or wild-type mice are maintained in SCF plus IL-3 *in vitro* (Fig. 1b). Moreover, limited

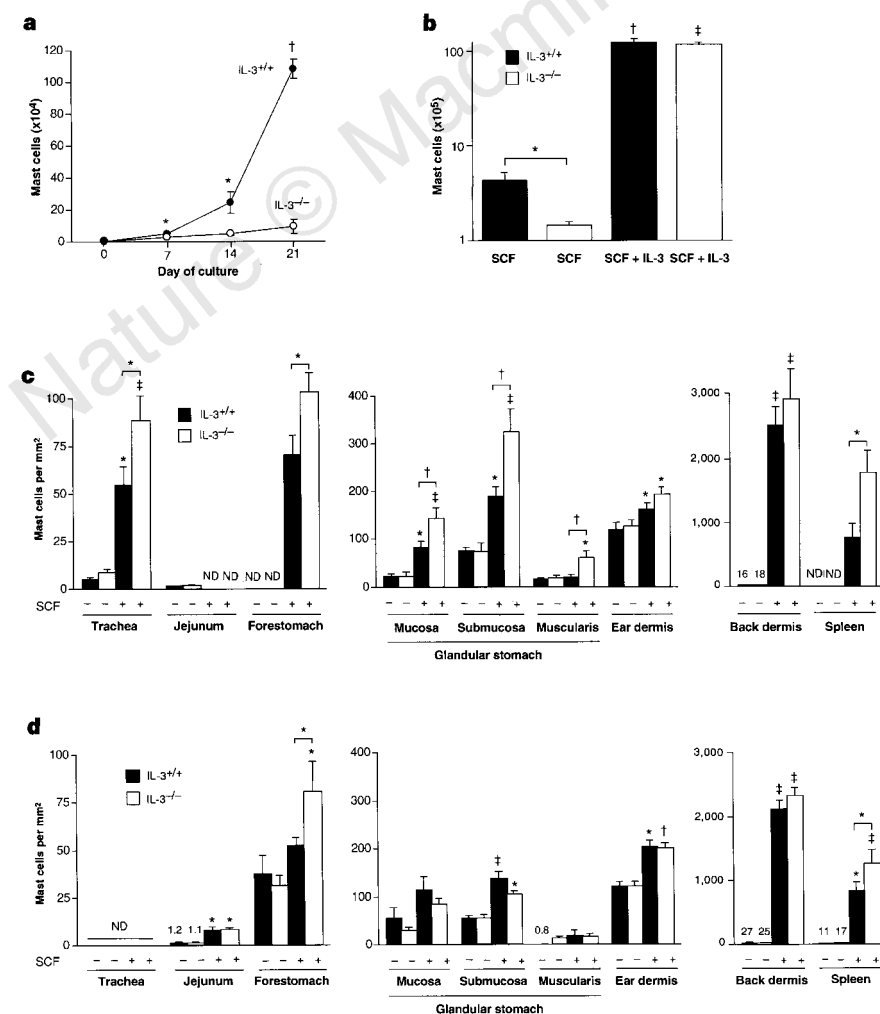


Figure 1 Differential requirements for IL-3 in SCF-induced mast-cell development. **a, b**, *In vitro*; **c, d**, *in vivo*. **a**, Development of mast cells from BALB/c IL-3^{-/-} mouse bone-marrow cells cultured *in vitro* with recombinant rat SCF (rrSCF) is markedly decreased compared with the development of mast cells from BALB/c IL-3^{+/+} mouse bone-marrow cells. All values are mean $\pm$ s.e.m. ($n = 3$) per 10⁶ bone-marrow cells plated and are representative of the results of 3 separate experiments. Similar results were obtained using 129Sv $\times$ C57BL/6 IL-3^{+/+} and IL-3^{-/-} mice. Asterisk, $P < 0.05$; dagger, $P < 0.001$ versus corresponding values for IL-3^{-/-} cells. **b**, Exogenous IL-3 enhances mast-cell development in cultures of BALB/c IL-3^{+/+} or IL-3^{-/-} bone-marrow cells maintained for two weeks in rrSCF. All results are mean $\pm$ s.e.m. ($n = 3-4$) per 10⁶ bone-marrow cells plated. Asterisk, $P < 0.05$; dagger, $P < 0.001$; double dagger, $P < 0.0001$ versus corresponding values for cells cultured with SCF alone or versus values indicated by the square bracket. **c, d**, Numbers of mast cells in various tissues of **(c)** rrSCF-treated or untreated 129Sv $\times$ C57BL/6 IL-3^{-/-} and IL-3^{+/+} mice or **(d)** rrSCF-treated or vehicle-treated BALB/c IL-3^{-/-} and IL-3^{+/+} mice. Mice were killed to assess mast-cell numbers at baseline **(c)** or 24 h after the last of 21 daily subcutaneous injections of rrSCF (100 μ g kg⁻¹ d⁻¹, **c, d**) or vehicle **(d)**. All results in **(c, d)** are expressed as mean $\pm$ s.e.m. ($n = 4-11$ mice per group). Statistical significance was determined by Student's *t*-test (two-tailed) or one-way analysis of variance. Asterisk, $P < 0.05$; dagger, $P < 0.001$; double dagger, $P < 0.0001$ versus corresponding values for untreated or vehicle-treated mice of the same genotype or (as indicated by square brackets) versus values for mice of the other genotype. ND, not done. +, plus SCF; -, vehicle-treated or untreated.

SCF-dependent mast-cell development can occur *in vitro* in the absence of exogenous or endogenous IL-3 (Fig. 1a, b).

To assess the role of IL-3 in mast-cell development *in vivo*, we counted mast cells in the tissues of IL-3^{-/-} and wild-type mice at baseline or after 21 daily subcutaneous injections of recombinant rat SCF (rrSCF, at 100 µg kg⁻¹ d⁻¹) or vehicle alone (Fig. 1c, d). The results of these experiments show that, *in vivo*, endogenous IL-3 is not essential either for the development of mast cells under physiological conditions or for rrSCF-induced mast-cell hyperplasia (overproduction). Indeed, in certain tissues, mast-cell levels after rrSCF-treatment were significantly greater (by up to 140%) in IL-3-deficient mice than in the corresponding wild-type mice (Fig. 1c, d).

We then counted mast cells and basophils, and assessed immunity to parasites, in IL-3^{-/-} and wild-type mice that had been infected with the intestinal nematode *Strongyloides venezuelensis*, which is a naturally occurring parasite of murine rodents. It is rejected by a T-cell-dependent immune response, which is associated with extensive mast-cell hyperplasia in the intestinal mucosa^{16,17}. In three separate experiments, we found that IL-3^{-/-} mice inoculated with 2,000 *S. venezuelensis* third-stage infective larvae (L3), in comparison to corresponding wild-type mice, exhibited both significantly delayed expulsion of the adult worms (data not shown) and significantly prolonged production of the parasite's eggs (Fig. 2a, b).

In addition, the IL-3^{-/-} mice that were infected with *S. venezuelensis* exhibited two striking abnormalities in their haematopoietic effector-cell response to the parasite. First, although baseline percentages of bone-marrow basophils were essentially identical in IL-3^{-/-} and wild-type mice, *S. venezuelensis* infection induced a significant increase in basophil levels in the wild-type mice but not

in the IL-3^{-/-} mice (Fig. 2c). These findings confirm the hypothesis, which had been based mainly on analyses of effects of recombinant IL-3 (refs 7, 8), that endogenous IL-3 can expand basophil populations *in vivo*.

Second, endogenous IL-3 was required for a substantial proportion (~76%), but not all, of the mast-cell hyperplasia that occurred in C57BL/6 mice near the main site of *S. venezuelensis* infection, the jejunum (Fig. 2d–f). IL-3 contributed less to jejunal mast-cell hyperplasia during *S. venezuelensis* infection of BALB/c mice (compare Fig. 2f with 2g), perhaps reflecting strain-dependent differences in levels of other cytokines that can influence mast-cell development in mice^{3–5,9,13–15}. In contrast, IL-3 was required for nearly all of the increases in the number of mast cells that developed in the ileum or spleen of *S. venezuelensis* infected C57BL/6 or BALB/c mice (Fig. 2f, g). In additional experiments with female IL-3^{-/-} or IL-3^{+/-} C57BL/6 mice, mice infected with 400 *S. venezuelensis* L3 showed no detectable egg production or changes in basophil percentages or mast-cell numbers, whereas the effects seen in mice infected with 10,000 *S. venezuelensis* L3 were similar to those shown in Fig. 2a, c, d, f for mice infected with 2,000 *S. venezuelensis* L3.

Host immunity to *S. venezuelensis* is also impaired in *Kit^W/Kit^{W-v}* mice¹⁷. Because the *c-kit* mutations in these mice result in markedly reduced *c-kit*/SCF signalling^{18,19}, *Kit^W/Kit^{W-v}* mice are profoundly mast-cell-deficient²⁰. But *Kit^W/Kit^{W-v}* mice have been reported to have normal levels of blood basophils²¹, as well as apparently adequate T-cell function²². Moreover, IL-3 can induce mast-cell development in *Kit^W/Kit^{W-v}* mice²³, which may partly account for the modest numbers of mast cells that develop in the intestines of

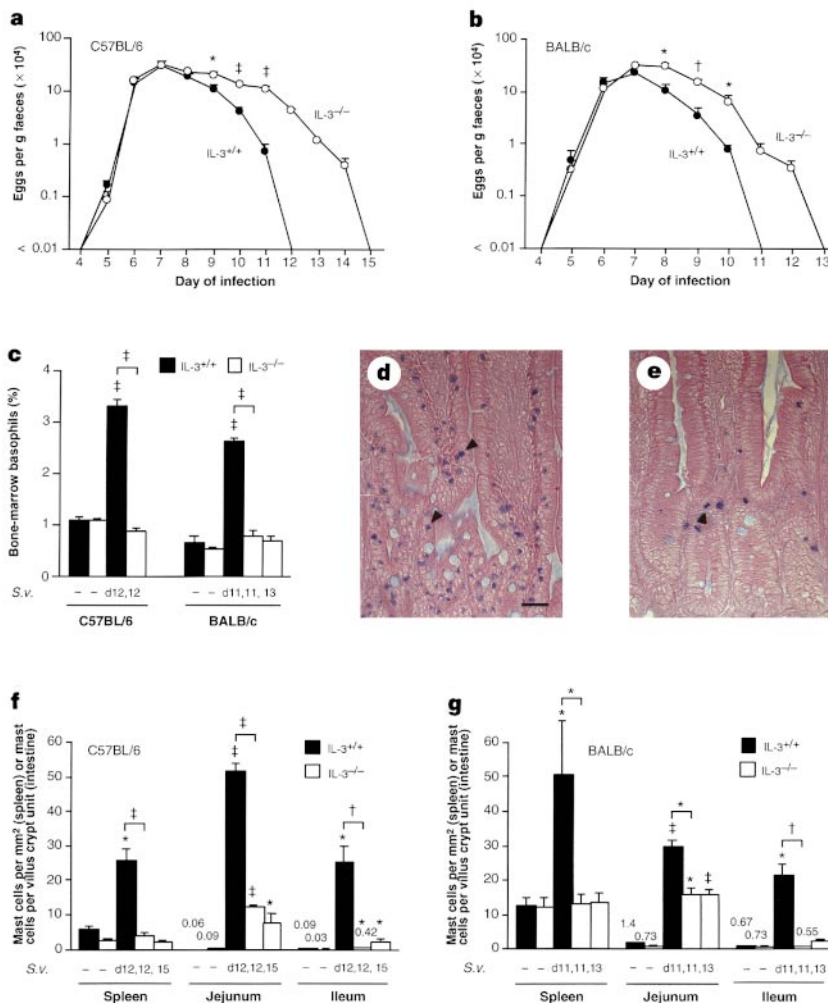


Figure 2 Defective parasite immunity and parasite-enhanced basophil and mast-cell development in IL-3^{-/-} mice compared with IL-3^{+/-} mice. **a, b**, Kinetics of *S. venezuelensis* (S.v.) egg production in male **(a)** C57BL/6 or **(b)** BALB/c IL-3^{-/-} and IL-3^{+/-} mice inoculated with 2,000 S.v. L3. All data are mean ± s.e.m. ($n = 4-11$ mice per point). Asterisk, $P < 0.05$; dagger, $P < 0.001$; double dagger, $P < 0.0001$ versus corresponding values for IL-3^{+/-} mice. Mice of any one given genotype in **a, b** all cleared the infection on the same day. Similar results were obtained in another experiment with C57BL/6 IL-3^{-/-} and IL-3^{+/-} mice. **c**, Percentage of bone-marrow basophils at baseline (–) or at various days (d) after S.v. infection in the IL-3^{+/-} and IL-3^{-/-} mice shown in **(a, b)**. All data are mean ± s.e.m. ($n = 3-6$ per group). Double dagger, $P < 0.0001$ versus corresponding baseline values or (as indicated by square brackets) versus corresponding values for mice of the other genotype. **d, e**, Mast cells (arrowheads) are much more abundant in the jejunal mucosa of S.v.-infected (d12) C57BL/6 IL-3^{+/-} mice **(d)** than in the corresponding tissue of the S.v.-infected (d12) C57BL/6 IL-3^{-/-} mice **(e)**. Scale bar in **d** represents 50 µm. **f, g**, Numbers of mast cells in the spleen, proximal jejunum and ileum at baseline (–) or at various days after S.v. infection in the IL-3^{+/-} and IL-3^{-/-} mice shown in **a, b**, respectively. All data are mean ± s.e.m., except that only mean values are given for very low values. ($n = 3-6$ per group). Asterisk, $P < 0.05$; dagger, $P < 0.001$; double dagger, $P < 0.0001$ versus corresponding baseline values or (as indicated by square brackets) versus corresponding values for mice of the other genotype.

these animals during infections with some parasites, including *S. venezuelensis*¹⁷. Finally, studies using a neutralizing antibody against SCF indicate that adequate SCF/c-kit signalling is required for the intestinal mast-cell hyperplasia induced by *Trichinella spiralis*, as well as for normal immunity to this helminth²⁴.

To examine *S. venezuelensis* infection of mice that have few mast cells and cannot make IL-3, we produced *Kit^W/Kit^{W-v}*, IL-3^{-/-} mice. Adult *Kit^W/Kit^{W-v}*, IL-3^{-/-} mice were clinically healthy and resembled *Kit^W/Kit^{W-v}*, IL-3^{+/+} mice in haematocrit and percentage of bone-marrow basophils at baseline (Fig. 3d). However, *Kit^W/Kit^{W-v}*, IL-3^{-/-} mice exhibited a more pronounced defect in their ability to reject *S. venezuelensis* than either *Kit^W/Kit^{W-v}*, IL-3^{+/+} mice (Fig. 3a, b) or *Kit^W/Kit^{W-v}*, IL-3^{-/-} mice (compare Fig. 3b with Fig. 2a, b). *Kit^W/Kit^{W-v}*, IL-3^{-/-} mice, unlike *Kit^W/Kit^{W-v}*, IL-3^{+/+} or wild-type mice, showed little or no enhancement of bone-marrow basophil production during *S. venezuelensis* infection (Fig. 3d). Moreover, *Kit^W/Kit^{W-v}*, IL-3^{-/-} mice infected with *S. venezuelensis* showed levels of histologically detectable mast cells in the jejunum, ileum and spleen that were even more reduced (compared with corresponding levels in wild-type mice) than those in the corresponding tissues in *S. venezuelensis*-infected *Kit^W/Kit^{W-v}*, IL-3^{+/+} mice (Fig. 3e, f).

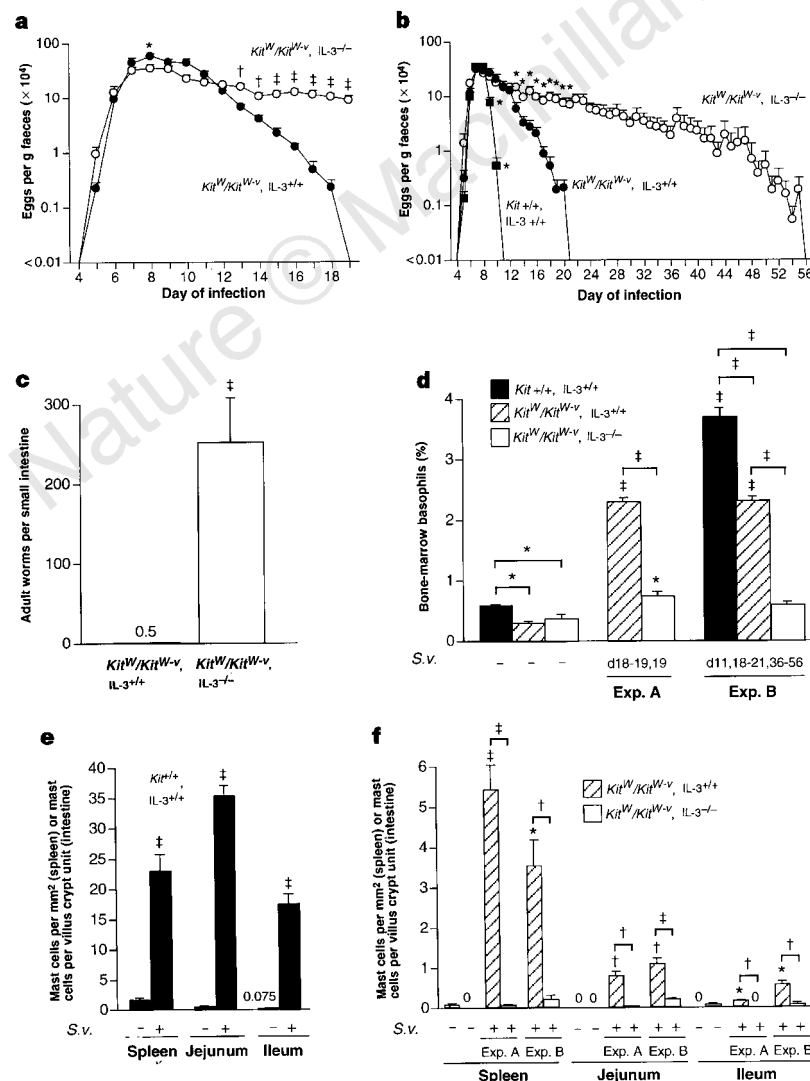
Abnormalities in addition to those affecting their mast-cell and basophil responses may have contributed to the delayed clearance of *S. venezuelensis* infections in IL-3^{-/-} or *Kit^W/Kit^{W-v}*, IL-3^{-/-} mice. The expression of contact hypersensitivity reactions (but not T-cell-dependent immunity to tumour cells) is moderately reduced in

IL-3^{-/-} mice¹⁰, indicating that IL-3^{-/-} mice may express defects in some T-cell-dependent responses that are not due solely to problems with mast-cell or basophil mobilization or function. In addition, *Kit^W/Kit^{W-v}* mice virtually lack interstitial cells of Cajal, which generate gut electrical pacemaker activity²⁵, and can exhibit reduced numbers of $\gamma\delta$ T cells in the gastrointestinal tract²⁶. Nevertheless, our findings clearly show that IL-3 contributes to the mast-cell hyperplasia and enhanced basophil development observed in mice during infection with *S. venezuelensis*. IL-3 is also needed for normal host immune responses to this nematode. Our data also indicate that IL-3 and SCF may have overlapping and/or synergistic roles in maintaining an adequate immune response to this parasite.

Note added in proof: We recently found that, compared with IL-3^{+/+} mice, IL-3^{-/-} mice infected with the parasite *Nippostrongylus brasiliensis* exhibited marked reductions in both hyperplasia of tissue mast cells and enhancement of bone-marrow basophil numbers, but no impairment of parasite expulsion. These findings suggest that the importance of IL-3 in immunity to parasites may vary according to the species of parasite. □

Methods

Generation of transgenic mice. IL-3^{-/-} mice¹⁰ were of the 129Sv × C57BL/6 background or were fourth backcross generation in a BALB/c or C57BL/6 background. Because *Kit^W/Kit^{W-v}* mice are sterile, *Kit^W/Kit^{W-v}*, IL-3^{-/-} mice were produced by first crossing C57BL/6 IL-3^{-/-} mice (fourth backcross generation)



with WBB6F1-*Kit*^{W/+} and *Kit*^{W-v/+} mice (Jackson Laboratory) to generate *Kit*^{W/+}, IL-3^{+/-} mice (50% of offspring) and *Kit*^{W-v/+}, IL-3^{+/-} mice (50%). These mice were bred with the IL-3^{-/-} mice to generate *Kit*^{W/+}, IL-3^{-/-} and *Kit*^{W-v/+}, IL-3^{-/-} mice (each 25% of offspring), which were bred to produce *Kit*^{W/Kit}^{W-v}, IL-3^{-/-} mice (25% of offspring). The IL-3 genotype of mice used for breeding was determined by Southern blotting (as in ref. 10), whereas the *c-kit* genotype was determined on the basis of coat colour and white-spotting appearance. All mouse experiments were conducted according to guidelines of the AAALAC-accredited BIDMC IACUC and NIH.

Bone-marrow culture. Femoral bone-marrow cells from individual mice (Fig. 1a) or pooled from five mice (Fig. 1b) 8–12 weeks of age were placed (initial density of 1×10^6 cells ml⁻¹) in DMEM containing 10% heat-inactivated fetal bovine serum, 50 μ M 2-mercaptoethanol, and 2 mM glutamine supplemented with rrSCF¹⁶⁴ (50 ng ml⁻¹, Amgen) $\pm$ IL-3 (50 units ml⁻¹, Genzyme) and maintained at 37 °C in a humidified atmosphere of 5% CO₂ in air²⁷. Non-adherent cells were transferred to new culture flasks containing 50% fresh medium one to two times per week and mast cells were counted at weekly intervals by examining cytocentrifuge slide preparations stained with Toluidine blue or May-Grünwald-Giemsa.

Treatment with SCF *in vivo*. Mice of both sexes (all at least six weeks old at the beginning of the experiment) received daily subcutaneous injections for 21 d (at the same back-skin site) of either polyethylene-glycol-conjugated rrSCF¹⁶⁴ (100 μ g kg⁻¹ d⁻¹; Amgen) or vehicle; 4 μ m paraffin sections of Carnoy's-fixed tissues were stained with safranin and 1.0% alcian blue at pH 1.0 (or, for intestines, pH 0.3) and mast cells were counted as the number per mm² of tissue (using the Bioquant Morphometric System, R & M Biometrics), or, for jejunum and ileum, as the number per villus crypt unit²⁸.

Infection with *S. venezuelensis*. *S. venezuelensis* were maintained by serial passage in male Wistar rats, and infectious L3 were obtained by faecal culture on filter paper²⁹. Mice (all at least six weeks old at the beginning of the experiment) were infected by subcutaneous inoculation with 400, 2,000 or 10,000 L3. The degree of infection of individual mice was monitored by counting the numbers both of eggs excreted daily (eggs per g faeces) and, at the time of killing, of adult intestinal worms. Bone-marrow basophils were identified and quantified by flow cytometry³⁰ and tissue mast cells were counted as described above.

Received 30th July; accepted 25th November 1997.

- Ihle, J. N. Interleukin-3 and hematopoiesis. *Chem. Immunol.* **51**, 65–106 (1992).
- Ihle, J. N. *et al.* Biological properties of homogeneous interleukin 3. I. Demonstration of WEHI-3 growth factor activity, mast cell growth factor activity, P cell stimulating factor activity and histamine producing factor activity. *J. Immunol.* **131**, 282–287 (1983).
- Madden, K. B. *et al.* Antibodies to IL-3 and IL-4 suppress helminth-induced intestinal mastocytosis. *J. Immunol.* **147**, 1387–1391 (1991).
- Finkelman, F. D. *et al.* Anti-cytokine antibodies as carrier proteins. Prolongation of *in vivo* effects of exogenous cytokines by injection of cytokine-anti-cytokine antibody complexes. *J. Immunol.* **151**, 1235–1244 (1993).
- Svetic, A. *et al.* A primary intestinal helminthic infection rapidly induces a gut-associated elevation of Th2-associated cytokines and IL-3. *J. Immunol.* **150**, 3434–3441 (1993).
- Rodewald, H.-R., Dessing, M., Dvorak, A. M. & Galli, S. J. Identification of a committed precursor for the mast cell lineage. *Science* **271**, 818–822 (1996).
- Saito, H. *et al.* Selective differentiation and proliferation of hematopoietic cells induced by recombinant human interleukins. *Proc. Natl Acad. Sci. USA* **85**, 2288–2292 (1988).
- Mayer, P., Valent, P., Schmidt, G., Liehl, E. & Bettelheim, P. The *in vivo* effects of recombinant human interleukin-3: demonstration of basophil differentiation factor, histamine-producing activity, and priming of GM-CSF-responsive progenitors in nonhuman primates. *Blood* **74**, 613–621 (1989).
- Paul, W. E., Seder, R. A. & Plaut, M. Lymphokine and cytokine production by Fc ϵ RI⁺ cells. *Adv. Immunol.* **53**, 1–29 (1993).
- Mach, N. *et al.* Involvement of interleukin-3 in delayed-type hypersensitivity. *Blood* (in the press).
- Ichihara, M. *et al.* Impaired interleukin-3 (IL-3) response of the A/J mouse is caused by a branch point deletion in the IL-3 receptor α subunit gene. *EMBO J.* **14**, 939–950 (1995).
- Nishinakamura, R., Miyajima, A., Mee, P. J., Tybulewicz, V. L. J. & Murray, R. Hematopoiesis in mice lacking the entire granulocyte-macrophage colony-stimulating factor/interleukin-3/interleukin-5 functions. *Blood* **88**, 2458–2464 (1996).
- Galli, S. J., Zsebo, K. M. & Geissler, E. N. The kit ligand, stem cell factor. *Adv. Immunol.* **55**, 1–96 (1994).
- Tsuji, K., Zsebo, K. M. & Ogawa, M. Murine mast cell colony formation supported by IL-3, IL-4, and recombinant rat stem cell factor, ligand for c-kit. *J. Cell. Physiol.* **148**, 362–369 (1991).
- Khalil, R. M. *et al.* *Schistosoma mansoni* infection in mice augments the capacity for interleukin 3 (IL-3) and IL-9 production and concurrently enlarges progenitor pools for mast cells and granulocytes-macrophages. *Infect. Immun.* **64**, 4960–4966 (1996).
- Nawa, Y. *et al.* Selective effector mechanisms for the expulsion of intestinal helminths. *Parasite Immunol.* **16**, 333–338 (1994).
- Khan, A. I., Horii, Y., Tiuria, R., Sato, Y. & Nawa, Y. Mucosal mast cells and the expulsive mechanisms of mice against *Strongyloides venezuelensis*. *Int. J. Parasitol.* **23**, 551–555 (1993).
- Nocka, K. *et al.* Molecular bases of dominant negative and loss of function mutations at the murine c-kit/white spotting locus: *W²⁷*, *W¹*, and *W*. *EMBO J.* **9**, 1805–1813 (1990).
- Hayashi, S., Kunisada, T., Ogawa, M., Yamaguchi, K. & Nishikawa, S. Exon skipping by mutation of an authentic splice site of c-kit gene in W/W mouse. *Nucleic Acids Res.* **19**, 1267–1271 (1991).

- Kitamura, Y., Go, S. & Hatanaka, K. Decrease of mast cells in W/W mice and their increase by bone marrow transplantation. *Blood* **52**, 447–452 (1978).
- Jacoby, W., Cammarata, P. V., Findlay, S. & Pincus, S. H. Anaphylaxis in mast cell-deficient mice. *J. Invest. Dermatol.* **83**, 302–304 (1984).
- Galli, S. J. & Kitamura, Y. Animal models of human disease. Genetically mast-cell-deficient W/W^u and S/S^u mice: their value for the analysis of the roles of mast cells in biological responses *in vivo*. *Am. J. Pathol.* **127**, 191–198 (1987).
- Ody, C., Kindler, V. & Vassalli, P. Interleukin 3 perfusion in W/W^u mice allows the development of macroscopic hematopoietic spleen colonies and restores cutaneous mast cell numbers. *J. Exp. Med.* **172**, 403–406 (1990).
- Donaldson, L. E., Schmitt, E., Huntley, J. F., Newlands, G. F. J. & Grecis, R. K. A critical role for stem cell factor and c-kit in host protective immunity to an intestinal helminth. *Int. Immunol.* **8**, 559–567 (1996).
- Maeda, H. *et al.* Requirement of c-kit for development of intestinal pacemaker system. *Development* **116**, 369–375 (1992).
- Puddington, L., Olson, S. & Lefrançois, L. Interactions between stem cell factor and c-kit are required for intestinal immune system homeostasis. *Immunity* **1**, 733–739 (1994).
- Tsai, M. *et al.* Induction of mast cell proliferation, maturation, and heparin synthesis by the rat c-kit ligand, stem cell factor. *Proc. Natl Acad. Sci. USA* **88**, 6382–6386 (1991).
- Tsai, M. *et al.* The rat c-kit ligand, stem cell factor, induces the development of connective tissue-type and mucosal mast cells *in vivo*. Analysis by anatomical distribution, histochemistry, and protease phenotype. *J. Exp. Med.* **174**, 125–131 (1991).
- Sato, Y. & Toma, H. Effects of spleen cells and serum on transfer of immunity to *Strongyloides venezuelensis* infection in hyothymic (nude) mice. *Int. J. Parasitol.* **20**, 63–67 (1990).
- Lantz, C. S. *et al.* IgE regulates mouse basophil Fc ϵ R1 expression *in vivo*. *J. Immunol.* **158**, 2517–2521 (1997).

Acknowledgements. We thank Z.-S. Wang, M. Yamaguchi, L. Fox and S. Fish for experimental assistance. This work was supported by USPHS grants (S.J.G.), the Claudia Adams Barr Foundation, a Young Markey Scientist Award and the Cancer Research Institute/Partridge Foundation (G.D.), the Roche Research Foundation (J.B.), the Korea Science and Engineering Foundation (C.H.S.) and the Swiss National Science Foundation (N.M.).

Correspondence and requests for materials should be addressed to S.J.G. (e-mail: sgalli@bidmc.harvard.edu).

Regulated nuclear import of Rel proteins in the *Drosophila* immune response

Louisa P. Wu & Kathryn V. Anderson

Molecular Biology Program, Memorial Sloan-Kettering Cancer Center and the Sloan-Kettering Division, Graduate School of Medical Sciences, Cornell University, 1275 York Avenue, New York, New York 10021, USA

The *Drosophila* immune response uses many of the same components as the mammalian innate immune response, including signalling pathways that activate transcription factors of the Rel/NK- κ B family^{1–4}. In response to infection, two Rel proteins, Dif and Dorsal, translocate from the cytoplasm to the nuclei of larval fat-body cells^{1,2,5}. The Toll signalling pathway, which controls dorsal-ventral patterning during *Drosophila* embryogenesis⁶, regulates the nuclear import of Dorsal in the immune response^{2,7}, but here we show that the Toll pathway is not required for nuclear import of Dif. Cytoplasmic retention of both Dorsal and Dif depends on Cactus protein; nuclear import of Dorsal and Dif is accompanied by degradation of Cactus. Therefore the two signalling pathways that target Cactus for degradation must discriminate between Cactus–Dorsal and Cactus–Dif complexes. We identified new genes that are required for normal induction of transcription of an antibacterial peptide during the immune response. Mutations in three of these genes prevent nuclear import of Dif in response to infection, and define new components of signalling pathways involving Rel. Mutations in three other genes cause constitutive nuclear localization of Dif; these mutations may block Rel protein activity by a novel mechanism.

Nuclear localization of Dorsal (Fig. 1) in the fat body depends on cytoplasmic components of the Toll signalling pathway², but the pathway that regulates Dif has not been defined. We found that Dif is properly translocated from fat-body cytoplasm to nuclei in response to infection in *Toll*^{-/-} and *pelle*^{-/-} larvae (Fig. 2a–c); the Toll pathway is therefore not required for the nuclear import of Dif in fat-body cells. The *cactus* gene encodes a member of the I κ B

family of proteins, which sequester Rel proteins in the cytoplasm before signalling^{8,9}. Loss of *cactus* causes constitutive nuclear localization of Dorsal in both embryos and the fat body^{2,10}. Dif is also constitutively nuclear in *cactus*^{-/-} fat-body cells (Fig. 2d–f). Therefore, the two signalling pathways that control Dif and Dorsal both use Cactus to retain the Rel proteins in the cytoplasm. Nuclear localization of Dorsal in the embryo is controlled by targeted degradation of the Cactus protein^{11–13}. Cactus is rapidly degraded in response to infection in wild-type larvae (data not shown), as well as *dorsal*^{-/-} larvae (Fig. 2g), indicating that nuclear localization of Dif is probably regulated through degradation of Cactus in a Cactus–Dif complex. This indicates that the upstream signalling machinery of one pathway may not recognize and degrade Cactus efficiently when Cactus is in a complex with the inappropriate Rel protein. In mammalian cells, phosphorylation of the specific serine residues of IκB that are required for ubiquitination and degradation of IκB occurs when IκB is found in a complex with NF-κB^{14–16}. Thus, mammalian signalling pathways may also discriminate between different Rel–IκB complexes and activate a specific Rel protein.

In a genetic screen for mutations on the third chromosome of *Drosophila* that interfere with the immune response (L.P.W. and K.V.A., unpublished observations), we identified a set of genes that are required to regulate the nuclear import of Rel proteins. During the *Drosophila* immune response, many antimicrobial peptides are secreted from the fat body into the haemolymph¹⁷. This response is mediated by a rapid induction of transcription of genes encoding these peptides. The Toll pathway (but not Dorsal) is important for activation of transcription of the antifungal gene *drosomycin*, but is not required for the activation of genes encoding antibacterial peptides such as dipterecin⁷; only a single gene, *immune deficiency* (*imd*), has been identified that is required for induction of dipterecin transcription^{7,18,19}. We identified 57 mutations in more than 40 different genes that prevent the induction of transcription of a reporter gene in which the dipterecin promoter was fused to the coding region of β-galactosidase (*dipt-lacZ*)²⁰ (see Methods). At 3 h after infection, wild-type larvae carrying the reporter gene had very high levels of β-galactosidase activity in fat-body cells, whereas larvae homozygous for any one of the 57 mutations did not stain (Fig. 3a). Analysis of larval RNAs confirmed that transcription of the endogenous dipterecin gene was not induced to normal levels in

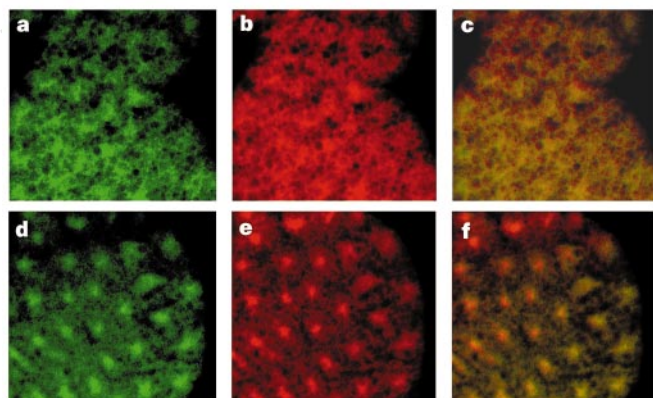


Figure 1 Dorsal and Dif nuclear localization in wild-type larval fat-body cells. **a–c**, In uninfected wild-type larvae, both Dif (green) (**a**) and Dorsal (red) (**b**) proteins are cytoplasmic. **c**, Superimposition of Dorsal and Dif labelling. Fat droplets in these cells give the cytoplasm a mesh-like appearance. **d–f**, Within 30 min after infection, both Dif (**d**) and Dorsal (**e**) are predominantly nuclear in localization. **f**, Superimposition of Dorsal and Dif labelling. The foci of Dif and Dorsal staining correspond to the positions of the cell nuclei, as seen by phase-contrast microscopy.

response to infection in the mutants (Fig. 3b). These genes were named *ird* (*immune response deficient*) genes (Table 1).

To investigate how the *ird* mutations interfered with dipterecin induction, we tested whether mutations in an arbitrary sample of six genes affected the subcellular localization of Dorsal and Dif before and after infection of homozygous mutant larvae. Surprisingly, all the mutations affected the nuclear localization of the Rel proteins. The *dorsal* and *Dif* genes are on the second chromosome, so any effect on Dorsal or Dif localization of mutations in genes on chromosome 3 must be a *trans* effect.

In larvae homozygous for mutations at three different loci, Dif was present in the fat-body cytoplasm before infection, but failed to move into nuclei in response to infection (compare Fig. 1d, Fig. 4a, d); we designated the mutants (*ird4*, *ird6* and *ird8*) that block nuclear localization of Dif class I mutants. In *ird4* and *ird8* mutants, the translocation of Dorsal from cytoplasm to nucleus in response to infection was normal in many cells (Fig. 4b), although there were patches of cells where Dorsal remained cytoplasmic (Table 1). Therefore, the *ird4* and *ird8* genes are required for the signal-dependent nuclear localization of Dif, and have a less important or indirect role in nuclear localization of Dorsal. The *18-wheeler* gene, which encodes a transmembrane protein that is homologous to Toll, is important for nuclear localization of Dif during the immune response²¹, so *ird4* and *ird8* may encode products that are necessary for *18-wheeler* activation or cytoplasmic components that act downstream of *18-wheeler*.

The *ird6* mutation prevents the nuclear localization of both Dif and Dorsal in response to infection, indicating that it may encode a component that is common to both Rel signalling pathways. The

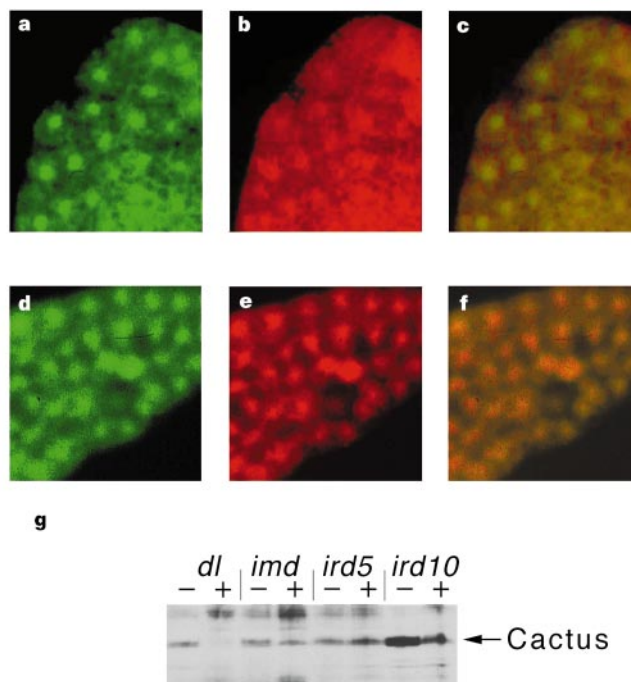


Figure 2 Toll is not required for nuclear localization of Dif. **a**, Dif moves to the nucleus in infected *Toll*^{-/-} larvae (*Df(3R)ro^{XB3}/Df(3R)7^{9QRX}*). **b**, Dorsal remains cytoplasmic in these larvae. **c**, Superimposition of Dif and Dorsal labelling. Similar results were obtained with *pelle*^{-/-} larvae (data not shown). **d–f**, Dif (**d**) and Dorsal (**e**) are both nuclear in uninfected *cactus*^{-/-} larvae (*cact^{HE9}/cact²⁵⁵* at 29°C). **f**, Superimposed images. **g**, The degradation of Cactus accompanies activation of Dif in *dorsal*^{-/-} larvae (*dl¹/dl¹*) (lanes 1, 2) (–, uninfected; +, 30 min after infection). Cactus protein is present in class II mutants, which cause constitutive nuclear localization of Dif (lanes 3–8).

2.2 kilobases of the *dipteracin* promoter used in the *dipt-lacZ* reporter gene include κ B, NF-IL6 and GATA-binding sites^{22,23}. No mutations in the *Dif* gene have been reported, but because the class I mutants were identified on the basis of decreased *dipt-lacZ* expression and because they block signal-dependent nuclear localization of Dif, nuclear translocation of Dif and binding of Dif to κ B-binding sites in the target promoter is likely to be an important inducible event that leads to *dipteracin* expression.

In class II mutants, *dipt-lacZ* is not induced and Dif is present in the fat-body nuclei both before and after infection (Fig. 5). Three of the six new genes examined are class II genes (*ird5*, *ird9* and *ird10*). We also found that mutations in the chromosome-2 gene *imd*, which do not induce transcription of *dipteracin*^{7,18,19} produce constitutively nuclear Dif localization (Fig. 5a–c). *imd* is therefore also a class II gene. As with the class I mutants, the class II mutants could be distinguished on the basis of their effects of Dorsal. Both Dorsal and Dif are constitutively nuclear in *imd* mutants, whereas Dorsal translocation in response to infection seems to be normal in *ird10*

mutants (Fig. 5d–f), and *ird5* and *ird9* mutants have some nuclear Dorsal in the absence of infection (Table 1).

The nature of the class II mutants shows that nuclear Dif and Dorsal are not sufficient for transcription of the *dipteracin* gene, indicating that other components are also required for induction of antimicrobial-gene expression. Furthermore, the class II mutants define additional steps that negatively regulate nuclear localization of the Rel proteins. Although constitutive degradation of Cactus could cause nuclear localization of the Rel proteins, the class II mutations do not affect steady-state levels of Cactus protein (Fig. 2g). Inactive nuclear NF- κ B in a complex with nuclear I κ B α has been seen in mammalian cells²⁴, but Cactus does not appear to be found in the nucleus in class II mutants (Fig. 5g, h). One hypothesis that fits these observations is that the class II genes are required to allow the formation of a nuclear complex of Dif with other proteins, and that this complex is required both for activation of *dipteracin* transcription and for turnover of the nuclear Rel proteins. An active mechanism that turns over nuclear Rel proteins has been seen in mouse embryo fibroblasts, where NF- κ B persists in the nucleus after stimulation when no I κ B α is present^{25,26}. Further studies of the class II genes may define additional components that regulate this process.

Further genetic analysis should order the functions of the class I and class II genes and define which genes mediate pathogen recognition and which act in the fat-body signalling pathways

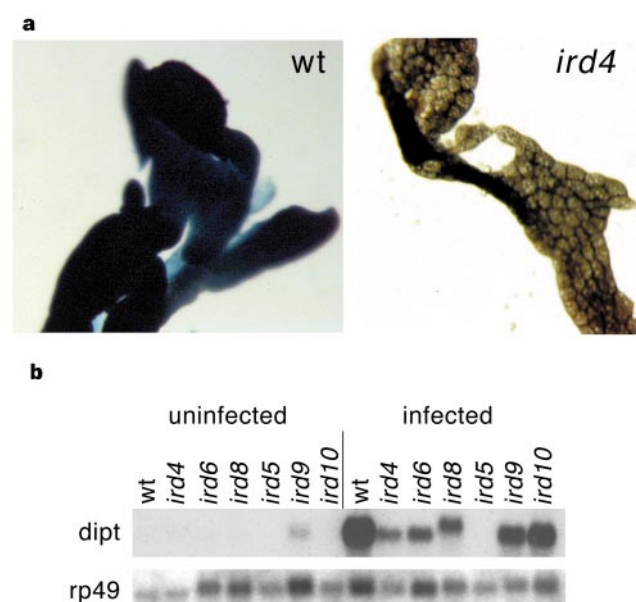


Figure 3 Mutants that affect *dipteracin* (*dipt*) expression. **a**, Fat bodies from wild-type and *ird4* mutants were stained for activity of the *dipteracin-lacZ* reporter gene after infection. Mutants were identified by the lack of staining. **b**, Northern blot showing the endogenous *dipteracin* messenger RNA before and after infection in wild-type and mutant larvae; the *rp49* mRNA is the loading control. Despite a clear difference in the induction of the reporter gene (**a**), the endogenous *dipteracin* gene (**b**) is induced to approximately normal levels in *ird9* and *ird10* mutants, indicating that the reporter gene may be particularly sensitive to the activity of Rel proteins.

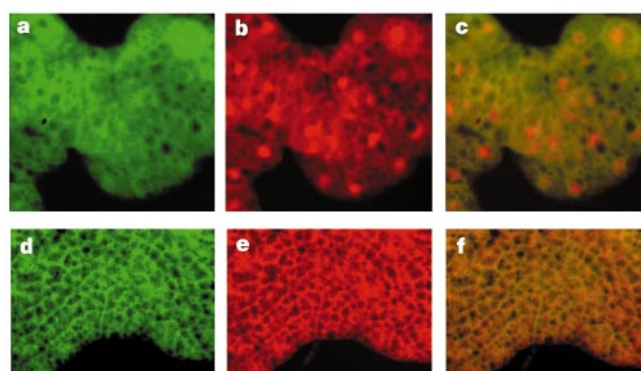


Figure 4 Class I mutants block nuclear localization of Dif and have variable effects on Dorsal. **a–c**, In larvae homozygous for the *ird4* mutation, Dif remains cytoplasmic in the fat body after infection (**a**), whereas Dorsal has moved from the cytoplasm to the nucleus (**b**); Dorsal and Dif staining is superimposed in **c**. **d–f**, In animals homozygous for the *ird6* mutation, both Dif (**d**) and Dorsal (**e**) remain cytoplasmic after infection; Dorsal and Dif staining is superimposed in **f**.

Table 1 Mutations that prevent *dipteracin* expression in response to infection

Gene	Map interval	Viability	Dorsal localization		Dif localization	
			Not infected	Infected	Not infected	Infected
Class I: Dif nuclear localization blocked						
<i>ird4</i>	<i>st-sr</i>	Lethal	C	C/N	C	C
<i>ird6</i>	<i>e-ca</i>	Lethal	C	C	C	C
<i>ird8</i>	<i>st-cu</i>	Viable	C	C/N	C	C
Class II: constitutive Dif nuclear localization						
<i>imd</i>	(chromosome II)	Viable	N	N	N	N
<i>ird5</i>	<i>cu-sr</i>	Semi-lethal	C/N	N	N	N
<i>ird9</i>	<i>cu-sr</i>	Viable	C/N	N	N	N
<i>ird10</i>	<i>st-cu</i>	Lethal	C	N	N	N

C, cytoplasmic; N, nuclear; C/N, patches of cells with nuclear Dorsal and other patches with cytoplasmic Dorsal in the fat body. In lethal lines, lethality is observed during pupal stages. Two alleles of *ird5* and *ird8* were identified in the screen; the other genes are represented by single alleles.

(Fig. 6). Studies of the activation of mammalian NF- κ B by the tumour-necrosis factor- α and interleukin-1 signalling pathways have identified a set of proteins that are required for Rel signalling, including components of the I κ B kinase complex and TRAF proteins, which have not been identified in *Drosophila*²⁷. As the *Drosophila* components that regulate nuclear localization of Dorsal were identified in maternal-effect screens that demand that mutant females survive to adulthood²⁵, such common signalling components could have been systematically overlooked in earlier genetic studies, but may be represented by the immune-response genes identified here. Given the apparent functional conservation of the insect and mammalian innate immune responses⁴, systematic

genetic analysis in *Drosophila* should provide important new views of the role of Rel proteins in innate immunity in all animals. □

Methods

Immunohistochemistry. Larvae were injected with a dilute solution of *Escherichia coli* and allowed to recover for 30 min before dissection. Larvae were dissected to expose the fat body, fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS) for 15 min, washed in PBS, then infiltrated with 1% Triton-X-100 and 1% BSA in PBS (PBT-A) for 30 min. The samples were then incubated with a 1:500 dilution of a rabbit anti-Dif antibody and a 1:10 dilution of a mouse monoclonal anti-Dorsal antibody in PBT-A for 1 h at room temperature. The mouse monoclonal anti-Cactus antibody was used at a 1:100 dilution. The samples were washed with 0.1% Triton-X-100 and 0.1% BSA in PBS (PBT-B), and then incubated in PBT-B with fluorescein isothiocyanate (FITC)-goat anti-rabbit-IgG serum and rhodamine-conjugated goat anti-mouse-IgG serum for 1 h. The tissue was then washed in PBT-B and the fat bodies were dissected out and mounted in 90% glycerol in PBS for microscopy.

Genetic screen. The screen will be described in detail elsewhere (L.P.W. and K.V.A., unpublished observations). In brief, homozygous *claret* males, carrying the *dipt-lacZ* reporter gene on chromosome 3, were mutagenized with 12.5 mM ethyl methane sulphonate. The mutagenized males were crossed with females carrying a dominant temperature-sensitive (*DTS4*) mutation over a balancer (*TM6B*) that carries the dominant larval marker *Tubby*. Single F₁ males were backcrossed to the same stock (*DTS4/TM6B*) and all the F₂ progeny carrying the *DTS4* chromosome were killed by shifting the crosses to 29 °C for 3 days. The remaining F₂ flies were transferred to new vials, and the F₃ flies that were homozygous for the mutagenized chromosome and for the *dipt-lacZ* reporter gene (the non-*Tubby* larvae) were screened for their immune responses.

To assay the induction of the *dipt-lacZ* reporter construct, five homozygous larvae from each line were injected with a dilute culture of *E. coli*. At 2.5 h after infection, the larvae were dissected open to expose the fat body and fixed in 1% glutaraldehyde in PBS. They were then washed and incubated in staining solution (0.2% X-gal in 0.01 M sodium phosphate buffer, pH 7.2, 0.15 M NaCl, 1 mM MgCl₂, 3.1 mM K₄[Fe^{II}(CN)₆], 3.1 mM K₃[Fe^{III}(CN)₆], 0.3% Triton-X-100) at 37 °C overnight. Of 6,429 lines that produced F₃ animals, 3,627 produced homozygous F₃ third-instar larvae; this rate of 44% lethal mutations indicates that the mutagenized third chromosomes contain an average of 0.58 new mutations that cause recessive lethality before late larval stages. Assuming that ~50% of all lethal mutations cause lethality before the third instar, there was an average of 1.2 lethal hits per chromosome, and the 3,627 lines tested should represent an average of almost two hits per gene. In an independent measure of the efficiency of mutagenesis, 11 mutations that inactivated the reporter gene itself were recovered. From the 3,627 lines tested, 57 lines reproducibly failed to induce the *dipt-lacZ* reporter gene. On the basis of a partial complementation matrix, we estimate that the screen identified mutations in roughly 40 different genes and that these represent 70–80% of the genes on the third chromosome that are required for normal induction of *dipt-lacZ*. The mutations in the six genes described here are the first lines to be characterized.

Northern and western analysis. Total RNA from third-instar larvae of the appropriate genotype, either in the absence of infection or 2 h after infection, was prepared using the Qiagen RNeasy kits; 4 μ g RNA was then blotted to Hybond nylon filters. Random-primed DNA probes (Boehringer) were made using the *dipterizin* and *rp49* complementary DNAs as templates. For western blots, four third-instar larvae were homogenized in 60 μ l SDS sample buffer and boiled for 5 min. Finally, 15 μ l was separated on a 7% polyacrylamide gel, and western blotting was carried out as described¹¹.

Received 6 November 1997; accepted 26 January 1998.

- Ip, Y. T. *et al.* Dif, a dorsal-related gene that mediates an immune response in *Drosophila*. *Cell* **75**, 753–763 (1993).
- Lemaître, B. *et al.* Functional analysis and regulation of nuclear import of Dorsal during the immune response in *Drosophila*. *EMBO J.* **14**, 536–545 (1995).
- Dushay, M., Åsling, B. & Hultmark, D. Origins of immunity: *Relish*, a compound Rel-like gene in the antibacterial defense of *Drosophila*. *Proc. Natl Acad. Sci. USA* **93**, 10343–10347 (1996).
- Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. Jr. A human homologue of *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997).
- Reichhart, J.-M. *et al.* Expression and nuclear translocation of the rel/NF- κ B-related morphogen dorsal during the immune response of *Drosophila*. *C. R. Acad. Sci. III* **316**, 1218–1224 (1993).
- Belvin, M. P. & Anderson, K. V. A conserved signaling pathway: The *Drosophila* Toll-Dorsal Pathway.

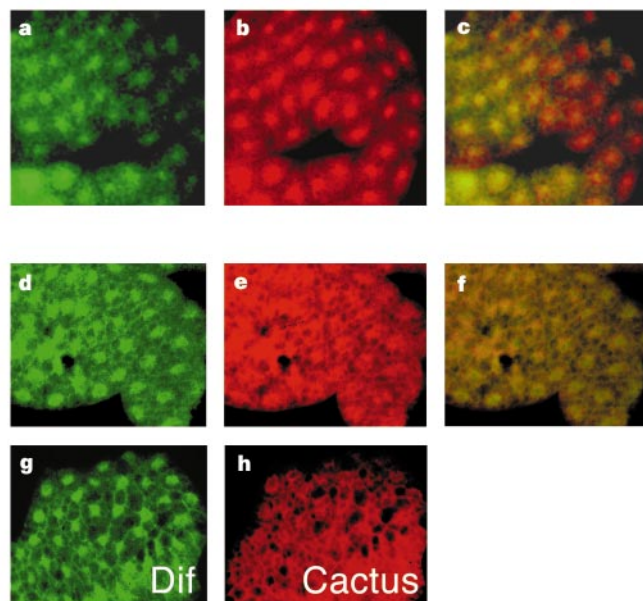


Figure 5 Class II mutants cause constitutive nuclear localization of Dif. **a–c**, In larvae homozygous for the *imd* mutation, both Dif (**a**) and Dorsal (**b**) are present in fat-body nuclei in the absence of infection; Dorsal and Dif staining is superimposed in **c**. The *imd* stock used is homozygous for the *imd Bc* chromosome¹⁸. **d–f**, In *ird10* homozygous mutant larvae, Dif is nuclear in the absence of infection (**d**), but Dorsal is predominantly cytoplasmic (**e**); Dorsal and Dif staining is superimposed in **f**. **g, h**, Cactus protein in uninfected *ird10* mutants is present in the cytoplasm (**h**) of the same cells in which Dif is nuclear (**g**).

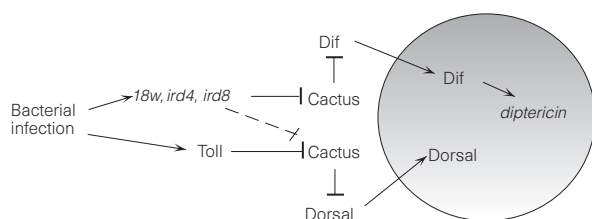


Figure 6 Signalling pathways that control the nuclear localization of Dif and Dorsal in the fat body in response to bacterial infection. The nucleus is shown by the shaded circle. Nuclear localization of Dorsal is activated by Toll, and nuclear localization of Dif requires *18-wheeler* (*18w*), *ird4* and *ird8*; both pathways act to inhibit Cactus by causing its degradation. The *ird6* gene is required for nuclear localization of both Dif and Dorsal, and may therefore act at an early step in pathogen recognition. The *ird4* and *ird8* mutants have some effect on the nuclear localization of Dorsal, suggesting that the *ird4/ird8*-dependent pathway also activates Dorsal at low efficiency. Dif may be required for transcription of *dipterizin*.

- Annu. Rev. Cell Dev. Biol.* **12**, 393–416 (1996).
7. Lemaître, B., Nicolas, E., Michaut, L., Reichardt, J.-M. & Hoffmann, J. A. The dorsoventral regulatory gene cassette *spätzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults. *Cell* **86**, 973–983 (1996).
 8. Geisler, R., Bergmann, A., Hiromi, Y. & Nüsslein-Volhard, C. *cactus*, a gene involved in dorsoventral pattern formation of *Drosophila*, is related to the I κ B gene family of vertebrates. *Cell* **71**, 613–621 (1992).
 9. Kidd, S. Characterization of the *Drosophila cactus* locus and analysis of interactions between cactus and dorsal proteins. *Cell* **71**, 623–635 (1992).
 10. Roth, S., Stein, D. & Nüsslein-Volhard, C. A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell* **59**, 1189–1202 (1989).
 11. Belvin, M. P., Jin, Y. & Anderson, K. V. Cactus protein degradation mediates *Drosophila* dorsal-ventral signaling. *Genes Dev.* **9**, 783–793 (1995).
 12. Bergmann, A. *et al.* A gradient of cytoplasmic Cactus degradation establishes the nuclear localization gradient of the dorsal morphogen in *Drosophila*. *Mech. Dev.* **60**, 109–123 (1996).
 13. Reach, M. *et al.* A gradient of Cactus protein degradation establishes dorsoventral polarity in the *Drosophila* embryo. *Dev. Biol.* **180**, 353–364 (1996).
 14. DiDonato, J. A., Mercurio, F. & Karin, M. Phosphorylation of I κ B precedes but is not sufficient for its dissociation from NF- κ B. *Mol. Cell Biol.* **15**, 1302–1311 (1995).
 15. Lin, Y. C., Brown, K. & Siebenlist, U. Activation of NF- κ B requires proteolysis of the inhibitor I κ B: signal-induced phosphorylation of I κ B alone does not release active NF- κ B. *Proc. Natl Acad. Sci. USA* **92**, 552–556 (1995).
 16. Miyamoto, S., Maki, M., Schmitt, M. J., Hatanaka, M. & Verma, I. M. Tumor necrosis factor α -induced phosphorylation of I κ B is a signal for its degradation but not dissociation from NF- κ B. *Proc. Natl Acad. Sci. USA* **91**, 12740–12744 (1994).
 17. Hoffmann, J. A. & Reichardt, J.-M. *Drosophila* immunity. *Trends Cell Biol.* **7**, 309–316 (1997).
 18. Lemaître, B. *et al.* A recessive mutation, *immune deficiency (imd)*, defines two distinct control pathways in the *Drosophila* host defense. *Proc. Natl Acad. Sci. USA* **92**, 9465–9469 (1995).
 19. Corbo, J. C. & Levine, M. Characterization of an immunodeficiency mutant in *Drosophila*. *Mech. Dev.* **55**, 211–220 (1996).
 20. Reichardt, J.-M. *et al.* Insect immunity: developmental and inducible activity of the *Drosophila* *dipterix* promoter. *EMBO J.* **11**, 1469–1477 (1992).
 21. Williams, M. J., Rodriguez, A., Kimbrell, D. A. & Eldon, E. D. The *18-wheeler* mutation reveals complex antibacterial gene regulation in *Drosophila* host defense. *EMBO J.* **16**, 6120–6130 (1997).
 22. Hoffmann, J. A., Hetru, C. & Reichardt, J.-M. The humoral antibacterial response of *Drosophila*. *FEBS Lett.* **325**, 63–66 (1993).
 23. Kadlaly, L., Petersen, U.-M. & Engström, Y. Adjacent GATA and κ B-like motifs regulate the expression of a *Drosophila* immune gene. *Nucleic Acids Res.* **25**, 1233–1239 (1997).
 24. Arenzana-Seisdedos, F. *et al.* Inducible nuclear expression of newly synthesized I κ B negatively regulates DNA-binding and transcriptional activities of NF- κ B. *Mol. Cell Biol.* **15**, 2689–2696 (1995).
 25. Beg, A. A., Sha, W. C., Bronson, R. T. & Baltimore, D. Constitutive NF- κ B activation, enhanced granulopoiesis and neonatal lethality in I κ B-deficient mice. *Genes Dev.* **9**, 2736–2746 (1995).
 26. Klement, J. F. *et al.* I κ B deficiency results in a sustained NF- κ B response and severe widespread dermatitis in mice. *Mol. Cell Biol.* **16**, 2341–2349 (1996).
 27. Stancovski, I. & Baltimore, D. NF- κ B activation: the I κ B kinase revealed? *Cell* **91**, 299–302 (1997).
 28. Anderson, K. V. & Nüsslein-Volhard, C. Information for the dorsal-ventral pattern of the *Drosophila* embryo is stored as maternal mRNA. *Nature* **311**, 223–227 (1984).

Acknowledgements. We thank R. Steward for antibodies to Dorsal and Cactus; Y. Engström for antibody against Dif; B. Lemaître for the *dipterix-lacZ* transgenic flies; M. Dushay for the *dipterix* cDNA; Y. Lu and S. Li for help with mapping; and M. Baylies, M. Belvin and T. Bestor for comments on the manuscript. This work was supported by NIH and NSF grants to K.V.A. and the Memorial Sloan-Kettering Cancer Center Support Grant. L.P.W. is a Leukemia Society of America Special Fellow.

Correspondence and requests for materials should be addressed to K.V.A. (e-mail: k-anderson@ski.mskcc.org).

The 1.7 Å crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven-bladed propeller

Louis Renault*†, Nicolas Nassar*§, Ingrid Vetter, Jörg Becker*, Christian Klebe*, Michel Roth† & Alfred Wittinghofer*

* Abteilung Strukturelle Biologie, Max-Planck-Institut für Molekulare Physiologie, Rheinlanddamm 201, 44139 Dortmund, Germany

† Institut de Biologie Structurale Jean-Pierre Ebel, Laboratoire de Cristallographie et de Cristallogénèse des Protéines, 41 Avenue des Martyrs, 38027 Grenoble cedex 1, France

The gene encoding the regulator of chromosome condensation (RCC1) was cloned by virtue of its ability to complement the temperature-sensitive phenotype of the hamster cell line tsBN2, which undergoes premature chromosome condensation or arrest in the G1 phase of the cell cycle at non-permissive temperatures^{1–2}.

‡ Permanent address: Institut de Biologie Structurale Jean-Pierre Ebel, Laboratoire de Cristallographie et de Cristallogénèse des Protéines.

§ Present address: Department of Chemistry, Baker Laboratory and Department of Pharmacology, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA.

RCC1 homologues have been identified in many eukaryotes, including budding and fission yeast. Mutations in the gene affect pre-messenger RNA processing and transport^{3,4}, mating⁵, initiation of mitosis⁶ and chromatin decondensation⁷, suggesting that RCC1 is important in the control of nucleo-cytoplasmic transport and the cell cycle. Biochemically, RCC1 is a guanine-nucleotide-exchange factor for the nuclear Ras homologue Ran⁸; it increases the dissociation of Ran-bound GDP by 10⁵-fold (ref. 9). It may also bind to DNA via a protein–protein complex². Here we show that the structure of human RCC1, solved to 1.7-Å resolution by X-ray crystallography, consists of a seven-bladed propeller formed from internal repeats of 51–68 residues per blade. The sequence and structure of the repeats differ from those of WD40-domain proteins, which also form seven-bladed propellers and include the β -subunits of G proteins. The nature of the structure explains the consequences of a wide range of known mutations. The region of the protein that is involved in guanine-nucleotide exchange is located opposite the region that is thought to be involved in chromosome binding.

The structure was solved by the single isomorphous replacement

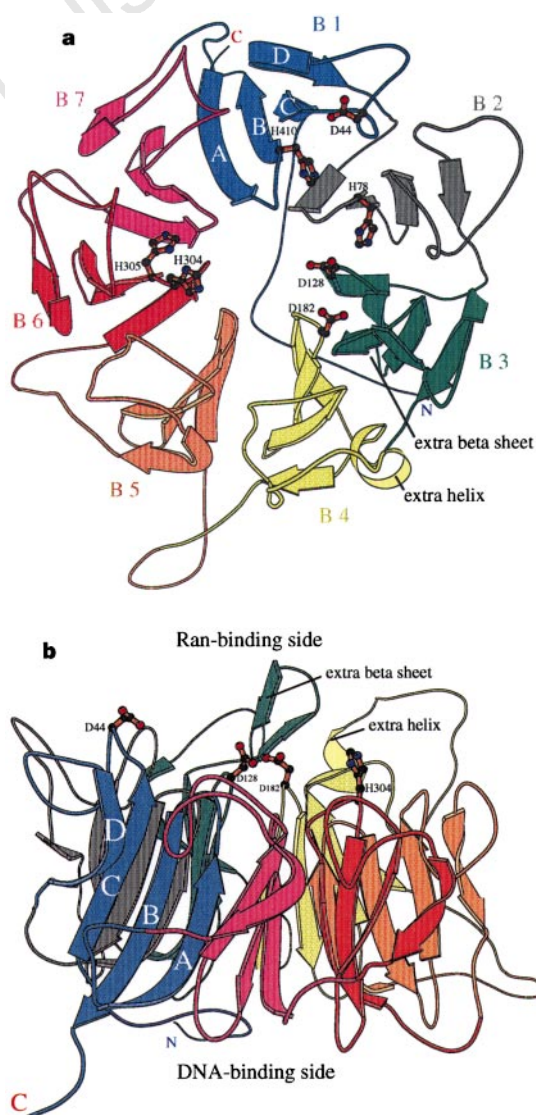


Figure 1 Overall three-dimensional structure of RCC1. Ribbon diagram of the RCC1 propeller structure as viewed along (a) or perpendicular to (b) the central shaft. The blades (a) are numbered (B1–B7) along the sequence. Semiconserved histidines that connect the blades and invariant residues believed to be important for the interaction with Ran are shown as ball-and-stick representations.

method (SIR) and non-crystallographic averaging. Briefly, a map calculated to 2.4 Å, with phases produced from one uranyl derivative, was subjected to iterative cycles of threefold non-crystallographic symmetry averaging and phase extension to 1.7 Å. The model consists of 400 amino acids (residues 21–421) per monomer and 886 water molecules per trimer. Residues 21–22, 232–239 and 419–421 are poorly defined. Final refinement statistics and statistics for the heavy-atom derivative are given in Table 1.

RCC1 consists mainly of β -strands. It has a pseudo-sevenfold symmetry with the overall appearance of a propeller made up of seven blades, each consisting of a four-stranded antiparallel β -sheet (Fig. 1a, b). The inner strand of the sheet is located along the shaft of the propeller. The amino- and carboxy-terminal tails lie on the same side of the propeller. The central shaft, which is filled with water molecules, has a diameter of 20 Å at one end (lower side, Fig. 1a) and narrows to 12 Å at the other (upper side). A similar propeller-like

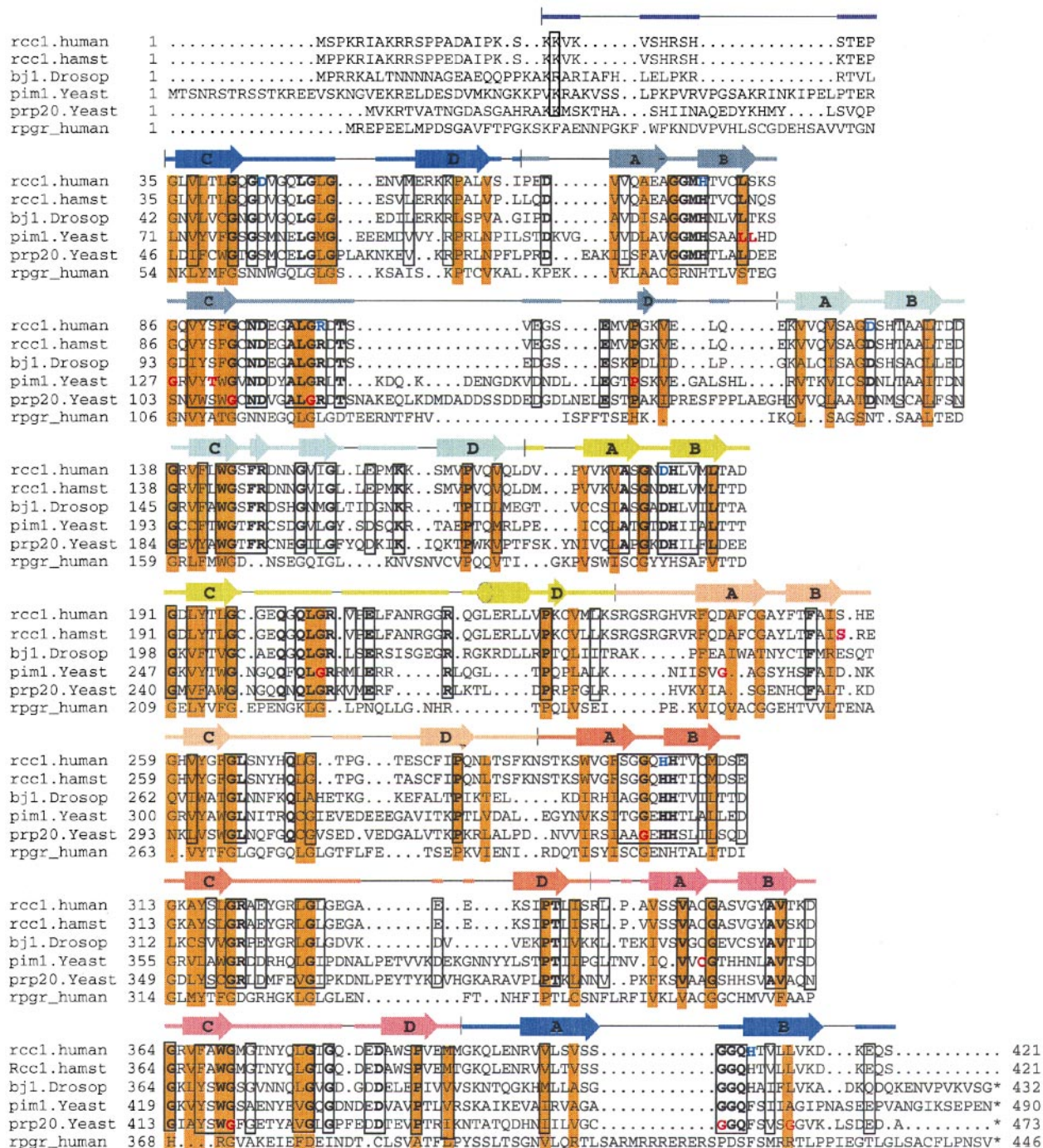


Figure 2 Primary- and secondary-structure alignment. Sequence alignment, obtained using the GCG program (adjusted manually) of five RCC1 homologues (from top to bottom: human, hamster, *Drosophila*, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*) and the protein product, RPGR/RP3 of the X-linked retinitis pigmentosa gene, along with the secondary structure of human RCC1. The seven blades are coloured as in Fig. 1. Residues highly conserved in seven RCC1 homologues (including *Xenopus* and *Caenorhabditis albicans*

RCC1) are boxed; invariant residues are also shown in bold. The structurally conserved residues among different repeats of RCC1 and RPGR/RP3 are shaded brown. Alanine mutations that perturb the GEF activity¹⁴ are shown in blue (human); the mutation responsible for the tsBN2 (refs 1, 2) phenotype is shown in pink (hamster); and RCC1 mutations in *S. cerevisiae* and *S. pombe*^{3-7,13} are shown in red. Asterisks indicate truncated sequences.

structure has also been found for the β -subunit of the heterotrimeric G proteins and for some enzymes^{10,11}. Sequence comparisons have shown that the β -subunits of G proteins belong to a large family of proteins that contain WD40 repeats¹². These repeats are ~40 residues in length and end in a conserved Trp–Asp motif (hence the name WD40).

An internal sevenfold sequence repeat of 51–68 residues for RCC1 (Fig. 2) has been observed earlier (ref. 1). However, the repeats identified from the structure (structural repeats) do not coincide with the sequence repeats, and are different from the WD40 repeats. In RCC1, the first half of the sequence repeat comprises the C and D strands of the structural repeat (Figs 1a and 2). Thus, one half of the first sequence repeat is made from the N-terminal end of the protein, and the other half is made from the C-terminal end. This arrangement seems to be needed to stabilize the circular arrangement of secondary structural elements, because one of the β -sheets is like a molecular clasp which tightens the circular belt of the ring structure. In the propeller structures of G β -subunits, this circular closure of the propeller is achieved by a 1 + 3 combination of β -strands from the N- and C-terminal protein ends.

An alignment of the seven repeats based on the three-dimensional structure shows seven highly conserved and six invariant residues. Four of the invariant residues are glycines and one is a *cis*-proline (Fig. 3a, b). The invariant glycines have ϕ/ψ angles that are not allowed for other amino acids and are sterically constrained, with no room to accommodate larger side chains. In addition to the residues that seem to be needed to be conserved to maintain the structural integrity of the propeller, there are several residues that are the same

in all RCC1 proteins. They are located in the loops (Fig. 1b) and are probably essential for the guanine-nucleotide-exchange activity of RCC1, for its ability to bind to chromosomes, and for other activities that it might have. As seen from a superimposition of the seven blades (Fig. 3b), the first three β -strands, which are close to the central tunnel, show less structural variation than the region on the outer side (around β -strand D). Most variation in length and structure is found in loops β C– β D and β D– β A, the latter one of which connects the different blades.

The RCC1 and WD40 repeats have different conserved residues. The only element that is similar to both types of propeller blade is the WD motif of the β -subunit and a motif of the RCC1 repeat, at the end of β -strand C, that contains a hydrophobic residue (often a W) and a glycine (the WG motif) (Figs 2, 3). Both of these motifs are involved in forming the hydrophobic core within the blades. In RCC1 the WG motif is part of a conserved VYxWG fingerprint motif, which characterizes each repeat.

Several budding yeast mutated genes encoding homologues of RCC1 have been identified. Screens for pheromone-response pathways have identified *srm* (for suppression of receptor mutations⁵); screens for mRNA splicing have identified *prp20* (for pre-RNA

Table 1 Summary of crystallographic analysis

Crystal Soaking conditions	Native*	Derivative Uranyl acetate saturated, 7.5 h	Derivative Uranyl acetate saturated, 15 h
Data collection			
Max. resolution (Å)	1.70	2.4	2.9
Completeness (%)†	94.3/79.9	86.8/27.6	64.5/16.7
Observed/unique‡	596,594/113,006	254,707/37,399	28,980/15,687
$\langle I/\sigma \rangle$ ‡	23.1/2.8	41.3/5.0	18.9/5.5
$R_{\text{sym}}^{\dagger\dagger}$ (%)‡	8.3/25.7	5.5/11.1	6.4/12.1
SIR phase determination			
Resolution range (Å)	20–2.4	20–2.4	20–2.9
$R_{\text{iso}}^{\dagger}$ (%)		19.7	18.1
Phasing power‡		2.20	2.90
$R_{\text{cullis}}^{\dagger}$		0.60	0.51
Number of sites#		7	7
Overall figure of merit	0.45		
Refinement			
Resolution (Å)	41–1.7		
$R_{\text{crist}}^{\dagger}$	18.7/26.0†	For 99,002 reflections	
$R_{\text{free}}^{\dagger\dagger}$ (%)	21.8/29.5	For 11,010 reflections	
Water molecules	886		
Ramachandran plot††	87.4	11.4	0.7
R.m.s. deviations‡‡	0.014	1.755	27.054
R.m.s.d. NCS protein (Å)§§	0.32	1.83	2.979

* Data set obtained by merging three native data sets (see Methods).

† All data/last shell. The last shells correspond to 1.73–1.7 Å for the native data set, and to 2.5–2.4 Å and 3.1–2.9 Å for the two derivative data sets.

‡ $R_{\text{sym}} = \sum_i \sum_j |I_i - I_j| / \sum_i \sum_j I_i$, where I_i is the measured intensity for reflection i , and $\langle I \rangle$ is the mean intensity.

§ $R_{\text{iso}} = \sum_i |F_{\text{PH}}| - |F_{\text{P}}| / \sum_i |F_{\text{PH}}|$, where F_{P} and F_{PH} are structure factors for native and derivative data.

|| Phasing power = $\langle F_{\text{Hcalc}} \rangle / \langle \text{Lack of closure} \rangle$, where $\langle F_{\text{Hcalc}} \rangle$ is the r.m.s. of the calculated heavy atom structure factors.

¶ $R_{\text{cullis}} = \langle \text{Lack of closure} \rangle / \langle \text{Isomorphous difference} \rangle$.

The weaker heavy-atom binding site (see Methods) involves the C groups of Glu 383 and Asp 382 of one monomer. Further refinement is required to confirm the interaction of stronger sites with solvent molecules and the backbone oxygen of Glu 56.

†† $R_{\text{crist}} = \sum_i |F_{\text{obs}}| - k |F_{\text{calc}}| / \sum_i |F_{\text{obs}}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors.

** R_{free} is the same as R_{crist} but calculated on 10% of the reflections that were always excluded from refinement.

††† Percentage of residues in most favoured, additionally allowed, generously allowed and disallowed regions, respectively, of the Ramachandran diagram.

‡‡ R.m.s. deviations for bond lengths (Å), bond angles (°), torsion angles (°) and improper torsion angles (°), respectively.

§§ Root mean squared distance (r.m.s.d.) for the core atoms (72.4% of the residues) and for all protein atoms related by the three-fold non-crystallographic symmetry (NCS), respectively.

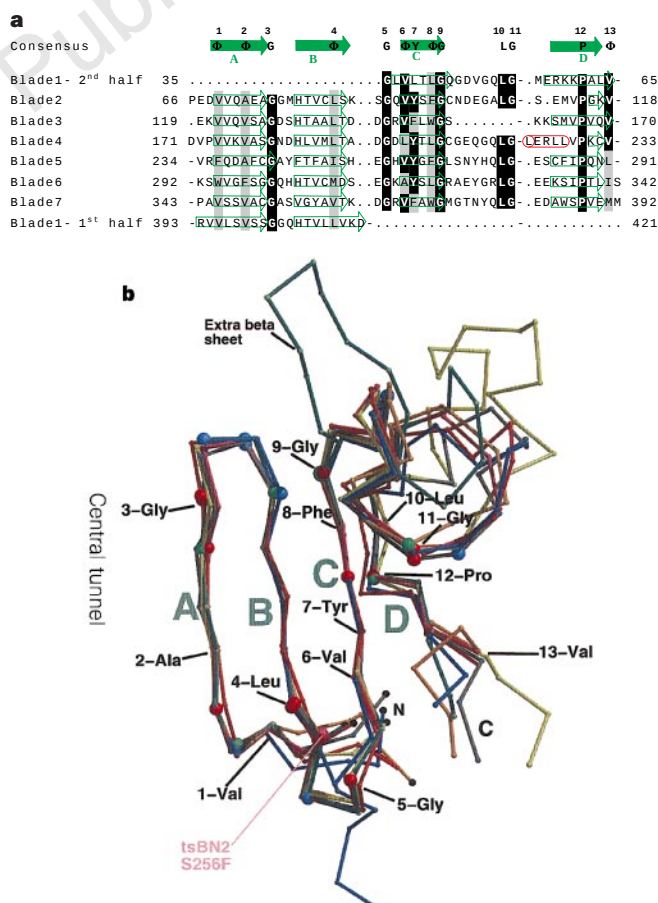


Figure 3 The structure of the RCC1 repeat. Sequence alignment and superimposition of the seven propeller blades, based on the structure. **a**, Primary- and secondary-structure alignment with the 13 invariant or highly conserved amino acids highlighted and numbered consecutively along the sequence. The top line shows the structural consensus, determined by common secondary-structure elements in five out of seven repeats. ϕ denotes any of the following hydrophobic residues: A, F, I, L, M, V or W. A dash indicates that the sequence is interrupted. Red residues mark the extra helix. **b**, Superimposition of the seven blades, indicating the structurally conserved residues (numbered from 1 to 13) and the variation in length. Coloured balls indicate the positions of mutations described and coloured identically in Fig. 2.

processing)³; and screens for mRNA transport have identified mtr-1 (ref. 4). In fission yeast, RCC1 has been isolated as pim1 (for premature initiation of mitosis), which is involved in initiation of mitosis⁶, as dcd (for defect in chromatin condensation), which is required for decondensation of mitotic chromosomes^{7,13}, and as sns (septated, not in S phase) mutants¹³. Most of these mutations (Figs 2, 3b) are temperature-sensitive, and affect residues conserved between the blades, indicating that these mutations may alter the architecture, rather than any specific interactions, of the protein. The nature of these mutations is often a glycine replaced by a charged or large hydrophobic group. As the invariant glycines have phi/psi angles disallowed for other residues and additionally steric constraints, these mutations appear to destabilize the structure at non-permissive temperatures.

The RCC1 gene of the hamster cell line tsBN2 carries a Ser 256 → Phe point mutation^{1,2} (Ser 256 of human RCC1). Ser 256 is located in a tight turn, connecting strands B and C of the fifth blade (Fig. 3b), where its side chain stabilizes the turn by making four potential hydrogen bonds with main-chain nitrogen atoms (not shown). Thus, the large hydrophobic side chain of phenylalanine at this position would reduce the stability of the protein.

RCC1 is the guanine-nucleotide-exchange factor (GEF) for Ran^{3,9}. An alanine scan of invariant, charged amino-acid residues has been used to identify residues of RCC1 that are involved in the GEF activity¹⁴. Mutations of Asp 44, His 78, Asp 128, Arg 101, Asp 182, His 304 and His 410 have been studied; these either reduce the affinity of the Ran–RCC1 interaction or have a significant effect on k_{cat} , the rate of the overall exchange reaction. In the three-dimensional structure, these residues are all located on the same side of the propeller (Fig. 1b) and thus could constitute the site of interaction with Ran.

Deletion of residues 8–29 of human RCC1 blocks the interaction with DNA^{1,2}, and, therefore, the DNA-interacting region is probably located opposite the Ran-interaction side. This conclusion is emphasized by mutation of Asp 136; proteins with this mutation are incapable of complementing the tsBN2 mutation. This mutation disrupts interactions with residue Lys 24 of the tail and might block interaction with DNA.

The interaction of a propeller scaffold with a GTP-binding protein seems to be evolutionarily conserved in two different (super)families of GTP-binding proteins, as shown by the RCC1–Ran interaction and the interaction of the β and α subunits of heterotrimeric G proteins^{10,11}. $G\alpha$ hovers over one side of the β -propeller and makes contacts with five of the seven blades. To locate the interface of Ran with the RCC1 propeller, we will need to determine the structure of the complex. However, propeller motifs seem to be well suited for the interaction between GTP-binding proteins and proteins that affect their nucleotide state. The structure of the heterotrimeric G protein stabilizes the binding of GDP, whereas the structure of the Ran–RCC1 complex reduces the affinity of Ran for GDP by 10^5 -fold and thus increases the rate of dissociation of nucleotide from the Ran–nucleotide–RCC1 ternary complex by the same order of magnitude⁹.

The RCC1-like repeat is encoded in the gene responsible for retinitis pigmentosa 3 (RP3), the major isoform of X-linked retinitis pigmentosa, xLRP^{15,16}. RCC1 and RP3 share homology (Fig. 2) and might have a similar function; therefore, RP3 is also known as retinitis pigmentosa GTPase regulator (RPGR). It has not been shown yet whether RPGR/RP3 is an exchange factor for Ran. As another locus associated with retinitis pigmentosa (more specifically, with the eye disease choroideremia) codes for the α subunit of the isoprenylating enzyme GGTase type II (ref. 17), RPGR/RP3 may be an exchange factor for a Rab protein. P532 is a large protein that consists of two RCC1 repeats and one WD40 repeat and thus combines three propeller domains¹⁸. P532 has guanine-nucleotide-exchange activity towards both Rab and Arf.

RCC1-like proteins might therefore be general guanine-nucleotide-exchange proteins.

The molecular mechanism by which a GEF catalyses nucleotide exchange on a Ras-related GTP-binding protein is unknown. The only structure available for a complex of a GTP-binding protein and its GEF is that of a complex of the elongation factors EF-Tu and EF-Ts^{19,20}. An insertion of residues of EF-Ts into the active site of EF-Tu disrupts the Mg-binding site¹⁹ or induces a peptide flip in the phosphate-binding loop²⁰ of EF-Tu. To view the Ran–RCC1 interaction, we have attempted to dock Ran onto the RCC1 propeller using the RCC1 mutations as a guideline. As nothing is known about RCC1-interacting sites on Ran, it will be vital to determine the three-dimensional structure of the Ran–RCC1 complex to understand the guanine-nucleotide-exchange mechanism more precisely. Determination of the structure of RCC1 is a first step towards this goal. □

Methods

Crystallization. Crystallization of human recombinant RCC1 (ref. 9), improvement of the poor quality crystals and the appropriate cryoconditions are described elsewhere (L.R., unpublished observations). Crystals diffract to 1.7 Å at –173 °C and belong to the space group *P1* ($a = 48.9$ Å, $b = 82.9$ Å, $c = 83.1$ Å, $\alpha = 112.9^\circ$, $\beta = 104.1^\circ$, $\gamma = 103.4^\circ$) with three molecules in the unit cell.

Data collection and processing. A high-resolution native data set at 1.7 Å resolution was collected at 100 K on a MAR image plate at the EMBL X11 beam-line (DESY, Hamburg) and processed with DENZO and SCALEPACK²¹. To improve the overall completeness and redundancy, this data set was merged with two native data sets previously collected on a rotating anode and processed with XDS²². Derivative data sets were collected from single crystals at 100 K on a rotating anode. We also attempted several multi-wavelength anomalous diffraction experiments with non-isomorphous crystals on the beam-line CRG D2AM at the ESRF synchrotron.

Structure solution and refinement. We analysed the uranyl acetate derivative data using the program suite CCP4²³. Native and derivative data sets were scaled with FHSCAL²³; difference Patterson maps then showed six strong heavy-atom sites that concurred with the non-crystallographic three-fold symmetry with two sites per monomer. The initial phases were calculated using MLPHARE²³ (Table 1) and refined by solvent flattening and histogram matching with DM²³. Addition of a second data set (Table 1) with the same heavy-atom binding sites improved the quality of the phases marginally. Refinement of the improper non-crystallographic three-fold-symmetry operators and construction of the first mask were performed with the density modification suite DEMON²⁴. Several cycles of density averaging (with DM) and reconstruction of the protein mask were employed then to build 90% of the model using the graphics program O (ref. 25). At this stage, a weaker heavy-atom site was found by self-difference Fourier. The model was refined using X-PLOR²⁶ with amplitude cut-off and with strict non-crystallographic symmetry (NCS) constraints until the R_{free} and R -factor reached 34.3% and 32.5%, respectively. Further refinement was carried out with NCS restraints and bulk-solvent corrections. The NCS restraints were released in the final stages of refinement, where automated refinement of protein models²⁷ was used with REFMAC²³. Use of this procedure led to an improved density for the residues of the flexible loops and for the residues that showed different conformations in the three monomers (residues 21–23, 168–172, 207–222, 232–239, 290–295 and 418–421). However, some of these amino acids (0.5% of the total number of amino acids) lie in the disallowed region of the Ramachandran plot.

All the figures showing parts of the model were drawn with Molscript²⁸; Fig. 3b was also drawn with Raster 3D²⁹.

Received 21 October; accepted 9 December 1997.

- Ohtsubo, M. *et al.* Isolation and characterization of the active cDNA of the human cell cycle gene (RCC1) involved in the regulation of onset of chromosome condensation. *Genes Dev.* **1**, 585–593 (1987).
- Seki, T., Hayashi, N. & Nishimoto, T. RCC1 in the Ran pathway. *J. Biochem.* **120**, 207–214 (1996).
- Aebi, M., Clark, M. W., Vijayraghavan, U. & Abelson, J. A yeast mutant, PRP20, altered in mRNA metabolism and maintenance of the nuclear structure, is defective in a gene homologous to the human gene RCC1 which is involved in the control of chromosome condensation. *J. Mol. Gen. Genet.* **224**, 72–80 (1990).
- Kadowaki, T., Goldfarb, D., Spitz, L. M., Tartakoff, A. M. & Ohno, M. Regulation of RNA processing

and transport by a nuclear guanine nucleotide release protein and members of the Ras superfamily. *EMBO J.* **12**, 2929–2937 (1993).

5. Clark, K. L. & Sprague, G. F. Jr Yeast pheromone response pathway: characterization of a suppressor that restores mating to receptorless mutants. *Mol. Cell Biol.* **9**, 2682–2694 (1989).
6. Matsumoto, T. & Beach, D. Premature initiation of mitosis in yeast lacking RCC1 or an interacting GTPase. *Cell* **66**, 347–360 (1991).
7. Sazer, S. & Nurse, P. A fission yeast RCC1-related protein is required for the mitosis to interphase transition. *EMBO J.* **13**, 606–615 (1994).
8. Bischoff, F. R. & Ponstingl, H. Catalysis of guanine nucleotide exchange on Ran by the mitotic regulator RCC1. *Nature* **354**, 80–82 (1991).
9. Klebe, C., Prinz, H., Wittinghofer, A. & Goody, R. S. The kinetic mechanism of Ran-nucleotide exchange catalysed by RCC1. *Biochemistry* **34**, 12543–12552 (1995).
10. Lambright, D. G. *et al.* The 2.0 Å crystal structure of a heterotrimeric G protein. *Nature* **379**, 311–319 (1996).
11. Wall, M. A. *et al.* The structure of the G protein heterotrimer $G_{\alpha 1}\beta_1\gamma_2$. *Cell* **83**, 1047–1058 (1995).
12. Neer, E. J., Schmidt, C. J., Nambudripad, R. & Smith, T. F. The ancient regulatory-protein family of WD-repeat proteins. *Nature* **371**, 297–300 (1994).
13. Matynia, A. *et al.* Isolation and characterization of fission yeast *sns* mutants defective at the mitosis to interphase transition. *Genetics* (in the press).
14. Azuma, Y. *et al.* Conserved histidine residues of RC1 are essential for nucleotide exchange on Ran. *J. Biochem.* **120**, 82–91 (1996).
15. Meindl, A. *et al.* A gene (RPRG) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). *Nature Genet.* **13**, 35–42 (1996).
16. Roepman, R. *et al.* Positional cloning of the gene for X-linked retinitis pigmentosa 3: homology with the guanine-nucleotide-exchange factor RCC1. *Hum. Mol. Genet.* **5**, 1035–1041 (1996).
17. Seabra, M. C., Brown, M. S. & Goldstein, J. L. Retinal degeneration in choroideremia: deficiency of rab geranylgeranyl transferase. *Science* **259**, 377–381 (1993).
18. Rosa, J. L., Casaroli-Marano, R. P., Buckler, A. J., Vilaró, S. & Barbacid, M. P532, a giant protein related to the chromosome condensation regulator RCC1, stimulates guanine nucleotide exchange on ARF1 and Rab proteins. *EMBO J.* **15**, 4262–4273 (1996).
19. Kawashima, T., Berthet-Colominas, C., Wulff, M., Cusack, S. & Leberman, R. The structure of the *Escherichia coli* EF-Tu-EF-Ts complex at 2.5 Å resolution. *Nature* **379**, 511–518 (1996).
20. Wang, Y., Jiang, Y., Meyerling-Voss, M., Sprinzl, M. & Sigler, P. B. Crystal structure of the EF-Tu-EF-Ts complex from *Thermus thermophilus*. *Nature Struct. Biol.* **4**, 650–656 (1997).
21. Otwinowski, Z. & Minor, W. In *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 556–562 (SERC, Warrington, UK, 1993).
22. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell contents. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
23. Collaborative Computational Project, N.4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
24. Vellieux, F. M. D., Hunt, J. E., Roy, S. & Read, R. J. DEMON/ANGEL: a suite of programs to carry out density modifications. *J. Appl. Crystallogr.* **28**, 347–351 (1995).
25. Jones, T. A. & Kjeldgaard, M. Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208 (1997).
26. Brünger, A. T. *X-PLOR Version 3.8* (Yale Univ., 1996).
27. Lamzin, V. S. & Wilson, K. S. Automated refinement of protein models. *Acta Crystallogr. D* **49**, 129–147 (1993).
28. Kraulis, P. J. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
29. Merritt, E. A., Murphy, M. E. P. Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869 (1994).

Acknowledgements. L.R. thanks the IBS/LCCP and the French Ministère de la Recherche et de l'Éducation for support and the staff at the European synchrotron radiation facility (ESRF) on the beam-line CRG D2AM for help. We thank the HFSP (A.W.); R. Schebaum for secretarial assistance; C. Nowak, N. Gülüm and M. Schulte-Spechtel for technical assistance; W. Kabsch and F. Vellieux for advice; the staff at the synchrotron facilities EMBL c/o DESY for support; B. Prakash for reading the manuscript and for discussions; and M. Hess for help with figures.

Correspondence and requests for materials should be addressed to A.W. (e-mail: alfred.wittinghofer@mpi-dortmund.mpg.de). Coordinates have been deposited at the Brookhaven Protein Data Bank, accession number A12.

Structure of the Sec7 domain of the Arf exchange factor ARNO

Jacqueline Cherfils, Julie Ménétrey, Magali Mathieu, Gérard Le Bras, Sylviane Robineau*, Sophie Béraud-Dufour*, Bruno Antonny* & Pierre Chardin*

Laboratoire d'Enzymologie et Biochimie Structurales, CNRS UPR 9063, 91198 Gif-sur-Yvette, France

* Institut de Pharmacologie Moléculaire et Cellulaire du CNRS-UPR 411, 660 Route des Lucioles, Sophia Antipolis, 06560 Valbonne, France

Small G proteins switch from a resting, GDP-bound state to an active, GTP-bound state. As spontaneous GDP release is slow, guanine-nucleotide-exchange factors (GEFs) are required to promote fast activation of small G proteins through replacement of GDP with GTP *in vivo*¹. Families of GEFs with no sequence similarity to other GEF families have now been assigned to most families of small G proteins. In the case of the small G protein Arf1, the exchange of bound GDP for GTP promotes the

Table 1 Crystallographic statistics

Data collection			
Crystal	Native	EMTS*, 1 mM	K ₂ HgI ₄ , 1 mM
Resolution limits	16–2 Å	16–3 Å	15–2.6 Å
R_{sym} † (%)	4.2	9.4	9.1
Observed reflections	70,772	27,275	110,828
Unique reflections	16,226	4,816	7,849
Completeness (%)	97.6	98.6	100
MIRAS analysis			
Number of sites		3	2
Phasing power Iso/Ano‡		2.25/1.77	2.16/1.39
R_{cullis} Iso/Ano§		0.66/0.77	0.71/0.78
Refinement			
Resolution	15–2 Å		
$R_{\text{factor}}/R_{\text{free}} $ (%)	18.2/22.6		
Average B (Å ²)	19.7		
Protein atoms	1,575		
Water molecules	142		
R.m.s.d. bond length¶ (Å)	0.009		
R.m.s. bond angle (°)	2.049		

* Derivative crystals were soaked for 1–2 weeks in heavy atom salts.

† $R_{\text{sym}} = \sum_i (I_i - \langle I \rangle) / \sum_i \langle I \rangle$ where I_i and $\langle I \rangle$ are the observed and average intensity for the reflection i .

‡ $R_{\text{cullis}} = (\text{phase-integrated lack of closure}) / (F_H - F_P)$ where F_H and F_P are the derivative and the protein structure factors. ISO, isomorphous; Ano, anomalous.

§ Phasing power = $(F_H / \text{phase-integrated lack of closure})$ where F_H is the calculated heavy atom structure factor.

|| $R_{\text{factor}} = \sum (F_o - F_c) / \sum F_o$ where F_o and F_c are the observed and calculated structure factor amplitudes. R_{free} is calculated as R_{factor} for a random set of reflections (10%) not included in the refinement.

¶ R.m.s.d., root mean square deviation from ideal values.

coating of secretory vesicles in Golgi traffic². An exchange factor for human Arf1, ARNO³, and two closely related proteins, named cytohesin 1 (ref. 4) and GPS1 (ref. 5), have been identified. These three proteins are modular proteins with an amino-terminal coiled-coil, a central Sec7-like domain and a carboxy-terminal pleckstrin homology domain. The Sec7 domain contains the exchange-factor activity³. It was first found in Sec7, a yeast protein involved in secretion⁶, and is present in several other proteins, including the yeast exchange factors for Arf, Gea1 and Gea2 (refs 7–9). Here we report the crystal structure of the Sec7 domain of human ARNO at 2 Å resolution and the identification of the site of interaction of ARNO with Arf.

The structure of the Sec7 domain of ARNO (Arf nucleotide-binding-site opener) was solved by X-ray diffraction, using multiple isomorphous replacement and anomalous scattering, and refined to an R_{factor} of 18.6% (Table 1). The protein is a flared cylinder of about 70 Å in length, with a diameter increasing from 20 Å at its N terminus to 40 Å at its C terminus (Fig. 1). It contains only helical secondary structures with 10 α -helices (labelled A to J), which stabilize the protein core mainly by hydrophobic and van der Waals interactions, with few hydrogen bonds. All helices and connecting loops have well-ordered electronic density except for residues 205–206 in loop H–I. Seven out of the ten loops feature one or two proline residues, all of which are conserved in most Sec7-homology sequences. The N-terminal helix extends by about three turns out of the protein core, but the rest of this helix packs against neighbouring helices. The helices are arranged in a distorted right-handed superhelix, which follows an axis from the N to the C terminus. A longer and more regular example of this superfold is found, for instance, in the Armadillo-repeat region of β -catenin¹⁰. No other all-helical structure found in the Protein Data Bank superimposes the fold of the ARNO Sec7 domain (ARNO-Sec7). ARNO-Sec7 bears no resemblance to the nucleotide-exchange factors EF-Ts^{11,12}, Mss4 (ref. 13) and GrpE¹⁴, nor to the exchange factor for Ran, regulator of chromosome condensation 1 (RCC1), which contains only β -sheets¹⁵. It is also unrelated to structures of other proteins that interact with G proteins, such as the GTPase-activating proteins for Ras¹⁶ and for Rho¹⁷.

Sec7 domains were previously defined by their primary structures

Full-length ARNO has an N-terminal coiled-coil domain involved in homodimerization (our unpublished results) and a C-terminal pleckstrin-homology (PH) domain^{3,4}. The N-terminal helix of ARNO-Sec7 protrudes from the globular core, forming an angle of $\sim 60^\circ$ with the axis of the superhelix, and could readily elongate into a coiled-coil. The C terminus of ARNO-Sec7, on the other hand, folds into a loop from Glu 241 to Pro 246 and forms

Sec7 domains feature two highly conserved sequences. 151 FRLPGE 156 and 187 VLSFAVIMLNTSLH 200 (ARN-Sec7 numbering), which will be called motif 1 and motif 2, respectively. Motif 1 forms the surface loop F–G, whose structure is stabilized by packing interactions with helix α D. It is largely accessible to the solvent, except for residues Phe 151 and Pro 154. Motif 2 folds into helix α H, with one face packed on helices α I and α J and the other exposed at the protein surface. Motif 1, motif 2 and several other conserved residues strikingly cluster on a concave region (Fig. 2). The groove is edged by the loop of motif 1 and the helix of motif 2, which face each other, with additional residues contributed by helix α F, helix α G and the C-terminal loop. The bottom of the groove forms a hydrophobic patch of ~ 15 Å in length and ~ 450 Å² in surface area. Polar and charged residues line its sides, providing

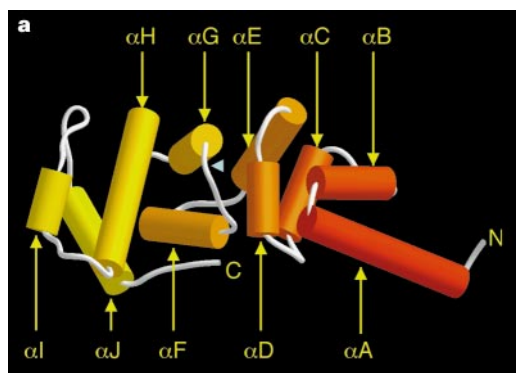
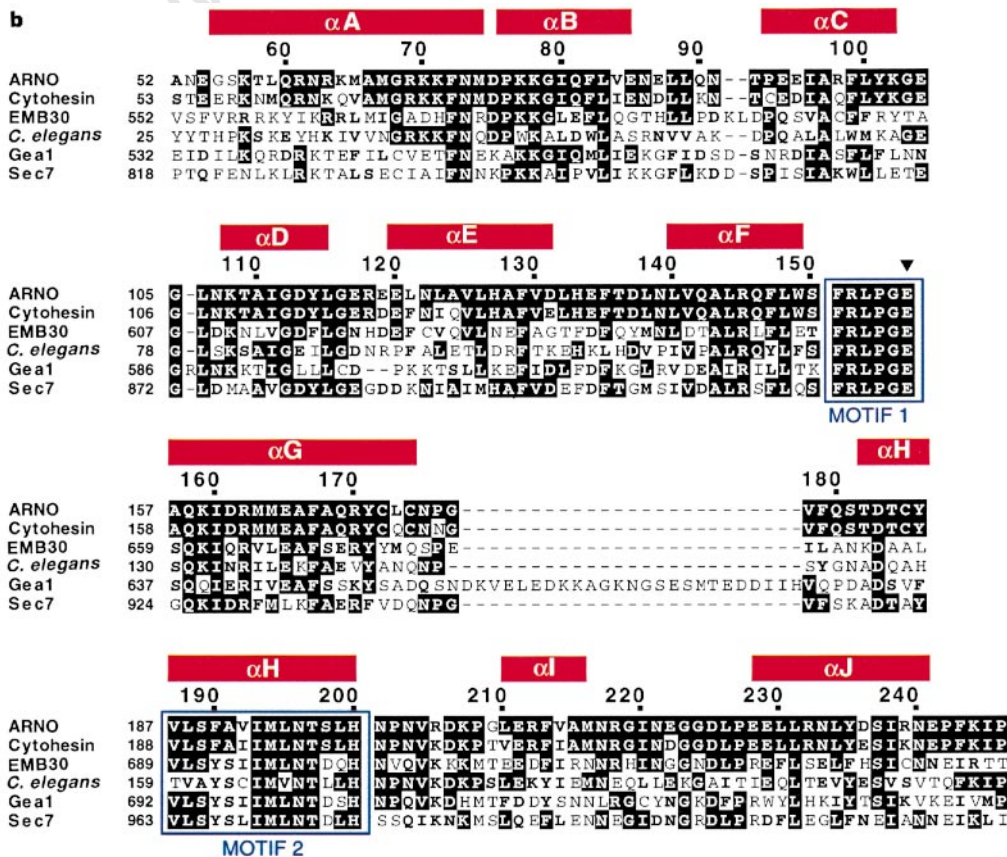


Figure 1 Structure of ARNO-Sec7. **a**, Overall view. The ten α -helices, numbered from A to J, are defined in Fig. 1b. They form a right-handed superhelix, coloured from red at the N terminus to yellow at the C terminus, whose axis is horizontal in the picture. Examples of this superfold are found in β -catenin¹⁰ and in a muramidase (Protein Data Bank entry 1SLY). The blue arrowhead marks the position of Glu 156 in loop F-G. **b**, Alignment of the ARNO-Sec7 sequence with sequences of homologous Sec7 domains from human cytohesin1, *Arabidopsis* EMB30, a *C. elegans* protein of unknown function, Gea1, a yeast exchange factor for Arf, and the secretion protein Sec7. The N- and C-terminal limits are as defined by the three-dimensional structure. The helices are shown in red boxes. Identical amino acids are boxed in black; conserved amino acids are shown in bold type. The two conserved motifs that form the putative Arf-binding site are boxed in blue. They fold into loop F-G and helix α H, respectively. The invariable residue Glu 156, which is involved in nucleotide exchange on Arf1, is indicated with a black arrowhead.



both donors and acceptors of hydrogen bonds (Fig. 3). The 'canyon' shape of this cluster of residues in ARNO-Sec7 and their high conservation in other sequences of Sec7 domains indicate that they may form the site for binding to Arf.

In *Arabidopsis* EMB30, charge reversal at the invariant Glu 658 (equivalent to Glu 156 in motif 1) produces severe defects in embryogenic differentiation⁸. To test the importance of this residue to nucleotide exchange on Arf1, Glu 156 was mutated to lysine in ARNO-Sec7. The mutant protein was unable to activate nucleotide exchange on myristoylated Arf1 (Fig. 4). Furthermore, structure-based mutagenesis of Arg 152 in motif 1 and of Met 194 and Asn 201 in motif 2 reduces the exchange activity by 5- to 100-fold (Fig. 4; S.B.-D. *et al.*, unpublished observations). In contrast, the mutation of Glu 117, which is further away from the groove, does not alter the nucleotide-exchange activity. Our structural and biochemical data are evidence that motifs 1 and 2 and the hydrophobic groove of ARNO define the binding site for Arf (Fig. 2b). The N-terminal coiled-coil domain in ARNO is far from the binding site for Arf1. On the other hand, the linker to the PH domain, and the PH domain itself, will be located next to the active site. However, binding of ARNO through its PH domain to vesicles containing phosphatidylinositol bisphosphate does not modify the intrinsic exchange activity of the Sec7 domain¹⁸.

Catalysis of nucleotide exchange on small G proteins by GEFs proceeds through the formation of a complex of the GEF and G protein; this complex stabilizes on release of the bound GDP^{11,12,15,19}. Then, GTP entry dissociates the complex. Such a tight 'transition state' complex has been seen between ARNO-Sec7 and the nucleotide-free form of a N-terminal truncated version of Arf1 (ref. 18).

The complex between nucleotide-free EF-Tu, a bacterial elongation factor with a guanine-nucleotide binding domain, and its exchange factor EF-Ts is, to date, the only model that describes the stereochemistry of nucleotide release from a G domain^{11,12}. EF-Ts wedges side chains in between the switch II region and its flanking helix, disrupting the coordination of the essential magnesium ion. Mutagenesis studies of p21^{ras} map the docking site for its exchange factors in regions equivalent to those on EF-Tu, including switch II (ref. 20) and helix $\alpha 3$ (ref. 21), indicating that the 'phosphate-side first' mechanism proposed for EF-Ts^{11,12} may operate for the exchange factors cdc25/SOS as well.

Arf1 is myristoylated on its N-terminal glycine and both the N-terminal helix and myristate are important for the nucleotide-exchange process in the presence of membranes²². ARNO was thus proposed to operate by binding to the myristate of Arf²³. However, ARNO as well as ARNO-Sec7 are effective on an N-terminal deletion mutant of Arf1 (ref. 18), ruling out an essential role for the fatty acid or the N-terminal helix in the catalytic process. In addition, mutation of Lys 73 into Glu in the switch II region of Arf1, which does not affect intrinsic GDP dissociation, abolishes the ability of ARNO to stimulate nucleotide exchange (data not shown), suggesting that switch II contributes to the binding site for ARNO. The switch II region of Arf1 (ref. 24) and other small G proteins has a flexible loop (residues 69–76 in Arf1) followed by helix $\alpha 2$ (77–85). Docking simulations using the crystal structure of unmyristoylated Arf1-GDP (ref. 24) indicate that the switch II region can be accommodated in the groove between motifs 1 and 2, with Arf burying a large surface area opposite its N-terminal helix. A model of interaction of the switch II of Arf with the groove of ARNO is reminiscent of the EF-Tu/EF-Ts complex and addresses the question of whether the exchange mechanism could be conserved among various members of the GTPase superfamily. However, as ARNO-Sec7 preferentially binds the nucleotide-free form of Arf¹⁸, which may adopt a different conformation as compared with Arf-GDP, the structure of the Arf1/ARNO-Sec7 complex will be needed for the detailed understanding of this interaction.

Nucleotide exchange on Ras and Rho proteins is mediated by protein modules of roughly the same size as the Sec7 domain, namely, the cdc25 and dbl domains, respectively (reviewed in ref. 1). These domains are predicted to be mostly helical according to secondary-structure-prediction algorithms, prompting us to investigate whether their primary structure is compatible with the all-helical fold found for ARNO-Sec7. A representative sequence of the cdc25 and dbl domains was threaded onto the structure of ARNO-Sec7 and compared with unrelated sequences²⁵. As neither cdc25 nor dbl emerged from the background, their overall structures may be unrelated to that of ARNO-Sec7. This does not rule out the possibility that they use similar catalytic mechanisms: this would not be unprecedented in regulation of small G proteins, as shown by the rasGAP and rhoGAP complexes^{16,17}. Both GAPs are mostly α -helical but with different folds; however, they use very similar mechanisms, positioning a catalytic arginine in the optimal conformation to promote nucleophilic attack of the γ -phosphate. The guanine-nucleotide-dissociation inhibitors for Rab²⁶ and Rho²⁷, too, have very different structures. Furthermore, the structure of the exchange factor for the small G protein Ran, RCC1, is a β -sheet propeller¹⁵ which bears no resemblance to the all- α -helical structure of ARNO-Sec7. It seems that although small G proteins have closely related structures, unrelated proteins have evolved to regulate them.

The structure of the Sec7 domain of ARNO has allowed us to locate the binding site for Arf and will provide a framework for further investigations of the nucleotide-exchange mechanism. In addition to functioning as a nucleotide exchange factor module for Arf, the Sec7 domain of the close homologue of ARNO, cytohesin 1, has been proposed to bind integrins and promote their adhesion to extracellular targets²⁸. It will be interesting to locate regions of the

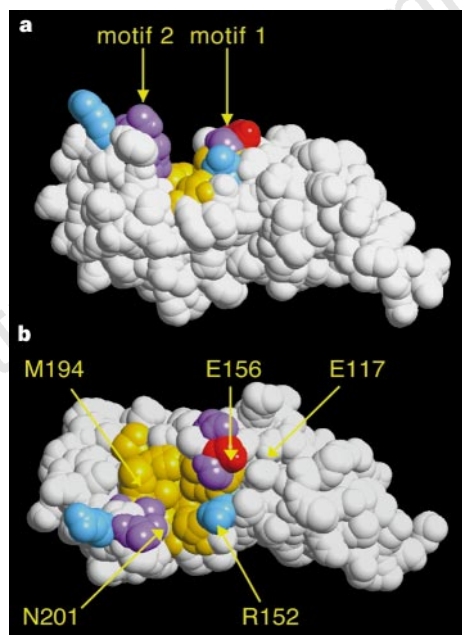


Figure 2 Topography of the conserved motives of ARNO-Sec7. **a**, **b**, The protein surface is shown in orthogonal views. The orientation in **b** is the same as in Fig. 1a and in **a** is rotated around a horizontal axis. Amino acids from motif 1 and motif 2 and conserved amino acids in contact with them form a connected surface, which has been coloured in yellow for hydrophobic residues, violet for polar residues, blue for basic residues and red for the acidic Glu 156. They define a groove at the protein surface; the bottom of this groove is hydrophobic and the sides are polar or charged. Residues from motif 1, motif 2, helix αF , helix αJ and the C terminus contribute to the hydrophobic patch. The two sides of the groove are from motif 1 and motif 2 respectively, with Glu 156 largely exposed to the solvent. Sites of single mutations (Fig. 4) are shown in **b**. Mutants of Arg 152, Glu 156, Met 194 or Asn 201 are defective for nucleotide exchange on Arf1. In contrast, mutation of Glu 117, outside the Arf1-binding site, has no effect.

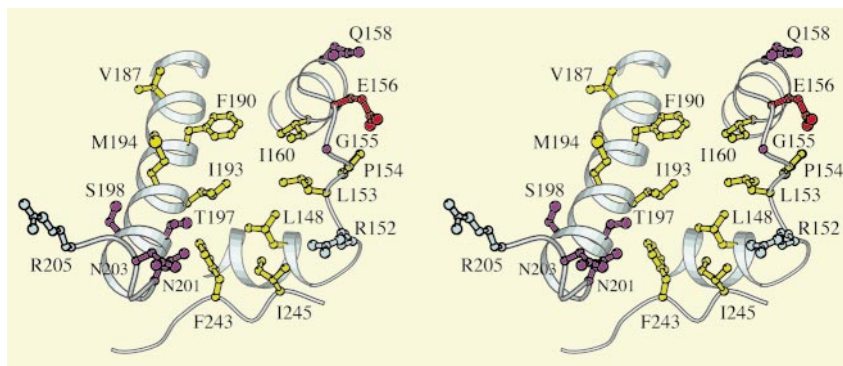


Figure 3 The putative site of binding to Arf. A close-up stereo view of Fig. 2b is shown, with the same orientation and colour-coding. Of the residues that compose the hydrophobic patch, Leu 148 in helix α F, Ile 160 in helix α G and Ile 193/Met 194 in motif2 are invariant among Sec7 domains; Val 187 in motif2 and Phe 243 in the C terminus are conserved; Phe 190 is generally replaced by tyrosine in other Sec7 domains and only Ile 245 in the C terminus is moderately

conserved. Acceptors and donors of hydrogen bonds in the presumed Arf-binding site are Arg 152/Glu 156 and the main chain NH of Gly 155 in motif 1, Gln 158 in helix α G, Thr 197/Ser 198 in motif2, and Gln 201/Gln 203/Arg 205 in loop H-I. Arg 152, Glu 156, Gln 158 and Thr 197 are invariant; the other amino acids are conserved. All the amino acids shown in this figure are accessible to the solvent and could contribute to the binding of Arf.

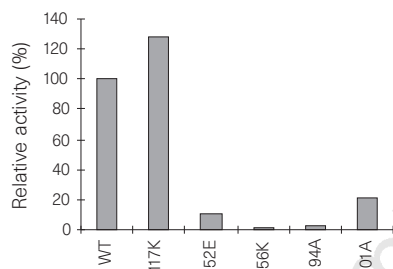


Figure 4 Nucleotide-exchange activity of ARNO-Sec7 mutants on Arf1. Catalysis of GDP-to-GTP exchange on myristoylated Arf1 by ARNO-Sec7 mutants was compared to the activity of wild-type ARNO-Sec7 (taken as a reference for 100% activity) as described in the Methods. All inactive mutants are from motif 1 and motif 2 on both sides of the hydrophobic groove (see Fig. 2). WT, wild-type.

implemented in SHARP³⁰. Subsequent identification of a third minor site in the EMTS derivative and of the two sites in the K₂HgI₄ derivative was carried out with the residual map facility in SHARP. Anomalous scattering was introduced in the heavy atoms parameter refinement at that stage for both derivatives. The ligands of the Hg atoms were found later to be the three cysteines of ARNO-Sec7, with the sites of the K₂HgI₄ derivative within 1 Å of the major sites of the EMTS derivative. Multiple isomorphous replacement combined with anomalous scattering (MIRAS) phases were improved by solvent flattening with SOLOMON²⁹ with a solvent content of 46%. The figure-of-merit weighted electron-density map at 10–2.6 Å resolution was easily interpretable at that stage, from which an initial model was built using program O. The initial *R* factor for that model was 38%. Experimental phases were not used later during refinement. Subsequent model building and positional refinement were carried out with TURBO and X-PLOR.

Site-directed mutagenesis and nucleotide-exchange assay. Mutants of ARNO-Sec7 were expressed in *E. coli* and purified as described for the wild-type protein³. Nucleotide exchange was measured by monitoring the intrinsic tryptophan fluorescence of Arf1 as described²². The assay was performed at 37 °C with 0.2 μM myristoylated Arf1-GDP, 0.1 μM ARNO-Sec7, 10 μM GTP and 0.4 mg ml⁻¹ unilamellar vesicles containing 70% phosphatidylcholine and 30% phosphatidylglycerol (S.B.-D. *et al.*, unpublished observations).

Structure analysis. Predictions of secondary structure were done with PHD (EMBL server www.embl-heidelberg.de). Docking of ARNO-Sec7 onto Arf1-GDP (ref. 24) (Protein Data Bank entry 1HUR) was performed manually and refined with an automated method based on surface complementarity of molecules treated as rigid entities. The cdc25 and dbl domain sequences threaded onto the ARNO-Sec7 structure were from the mouse exchange factor for Ras, cdc25, and the human exchange factor for Rho, Dbl, respectively. A hundred unrelated sequences were threaded onto ARNO-Sec7 to provide a statistical background²⁵.

Sec7 domain that might be involved in this process or in other protein–protein interactions. □

Methods

Crystallization. A truncated form of human ARNO (residues 51–252) encompassing the Sec7-homology domain was overexpressed in *Escherichia coli* and purified as described³. Crystals were grown in a hanging drop formed by a 1:1 mixture of the protein at 8 mg ml⁻¹ and the reservoir crystallization solution, containing 12.5% PEG 4000, 200 mM LiSO₄, 3.5% propanol and 1 mM DTT buffered in 100 mM EPPS at pH 8.5. Stabilization in cryoprotectant conditions was achieved by increasing the concentration of PEG 4000 and adding 20% glycerol. Heavy-atom derivatives were prepared in the stabilization solution supplemented with heavy-atom salts. The crystals belong to space group *P*2₁3 (*a* = *b* = *c* = 89.78 Å, α = β = γ = 90°), with one molecule in the asymmetric unit.

Data collection and processing. The native and EMTS (ethyl mercury thiosalicylate) derivative data sets were collected at 4 °C on a Mar image plate at beam line W32 of the LURE synchrotron with λ = 0.97 Å (Orsay). The K₂HgI₄ derivative data were recorded at the λ_2 absorption edge of Hg (1.0079 Å) at 100 K on a charge-coupled device (CCD) camera at beam line D2AM of the ESRF synchrotron (Grenoble). CCD images were converted to pseudo Mar images with IMAC (written by M. Roth). Data were processed with DENZO and scaled with SCALEPACK.

Phasing and refinement. Two heavy-atom sites were first located for the EMTS derivative from isomorphous difference Patterson maps using RSPS²⁹, and their parameters were refined with the maximum likelihood method

Received 21 November 1997; accepted 22 January 1998.

- Quilliam, L. A., Khosravi-Far, R., Huff, S. Y., Solski, P. A. & Der, C. J. Guanine nucleotide exchange factors: activators of the Ras superfamily of proteins. *Bioessays* **17**, 395–404 (1995).
- Scheckman, R. & Orci, L. Coat proteins and vesicle budding. *Science* **271**, 1526–1533 (1996).
- Chardin, P. *et al.* A human exchange factor for ARF contains Sec7- and pleckstrin-homology domains. *Nature* **384**, 481–484 (1996).
- Meacci, E., Tsai, S. C., Adamik, R., Moss, J. & Vaughan, M. Cytohesin-1, a cytosolic guanine nucleotide-exchange protein for ADP-ribosylation factor. *Proc. Natl Acad. Sci.* **94**, 1745–1748 (1997).
- Klarlund, J. K. *et al.* Signaling by phosphoinositide-3,4,5-trisphosphate through proteins containing pleckstrin and Sec7 homology domains. *Science* **275**, 1927–1930 (1997).
- Novick, P., Ferro, S. & Schekman, R. Orders of events in the yeast secretory pathway. *Cell* **25**, 461–469 (1981).
- Peyroche, A., Paris, S. & Jackson, C. L. Nucleotide exchange on ARF mediated by yeast Gea1 protein. *Nature* **384**, 479–481 (1996).
- Shevell, D. E. *et al.* EMB30 is essential for normal cell division, cell expansion, and cell adhesion in *Arabidopsis* and encodes a protein that has similarity to Sec7. *Cell* **77**, 1051–1062 (1994).
- Morinaga, N., Moss, J. & Vaughan, M. Cloning and expression of a cDNA encoding a bovine brain brefeldin A-sensitive guanine nucleotide-exchange protein for ADP-ribosylation factor. *Proc. Natl Acad. Sci. USA* **94**, 12926–12931 (1997).

10. Huber, A. H., Nelson, J. & Wies, W. I. Three-dimensional structure of the Armadillo repeat of β -catenin. *Cell* **90**, 871–882 (1997).
11. Kawashima, T., Berthet-Colominas, C., Wulff, M., Cusack, S. & Leberman, R. The structure of the *Escherichia coli* EF-Tu.EF-Ts complex at 2.5 Å resolution. *Nature* **379**, 511–518 (1996).
12. Wang, Y., Jiang, Y., Meyering-Voss, M., Sprinzl, M. & Sigler, P. Crystal structure of the EF-Tu.EF-Ts complex from *Thermus thermophilus*. *Nature Struct. Biol.* **4**, 650–656 (1997).
13. Yu, H. & Schreiber, S. L. Structure of guanine-nucleotide-exchange factor human Mss4 and identification of its Rab-interacting surface. *Nature* **376**, 788–791 (1995).
14. Harrison, C. J., Hayer-Hartl, M., Di Liberto, M., Hartl, F. U. & Kuriyan, J. Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. *Science* **276**, 431–435 (1997).
15. Renault, L. *et al.* The 1.7 Å crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven-bladed propeller. *Nature* **392**, 97–101 (1998).
16. Scheffzek, K. *et al.* The ras-rasGAP complex: structural basis for GTPase activation and its loss in oncogenic mutants. *Science* **277**, 333–338 (1997).
17. Rittinger, K., Walker, P. A., Eccleston, J. F., Smerdon, S. J. & Gamblin, S. J. Structure at 1.65 Å of RhoA and its GTPase-activating protein in complex with a transition-state analog. *Nature* **389**, 758–762 (1997).
18. Paris, S. *et al.* Role of protein-phospholipid interactions in the activation of ARF1 by the guanine nucleotide exchange factor Arno. *J. Biol. Chem.* **272**, 22221–22226 (1997).
19. Jacquet, E., Baouz, S. & Parmeggiani, A. Characterization of a mammalian C-CDC25Mm exchange factor and kinetic properties of the exchange reaction intermediate p21.C-CDC25Mm. *Biochemistry* **34**, 12347–12354 (1995).
20. Verroti, A. C. *et al.* RAS residues that are distant from the GDP binding site play a critical role in dissociation factor-stimulated release of GDP. *EMBO J.* **11**, 2855–2862 (1992).
21. Segal, M., Willumsen, B. & Levitzky, A. Residues crucial for Ras interaction with GDP-GTP exchangers. *Proc. Natl Acad. Sci. USA* **90**, 5564–5568 (1993).
22. Antony, B., Béraud-Dufour, S., Chardin, P. & Chabre, M. N-terminal hydrophobic residues of the G-protein ADP-ribosylation factor-1 insert into membrane phospholipids upon GDP to GTP exchange. *Biochemistry* **36**, 4675–4684 (1997).
23. Schimmöller, F., Itin, C. & Pfeffer, S. Vesicle traffic: get your coat! *Curr. Biol.* **7**, R235–R237 (1997).
24. Amor, J. C., Harrison, D. H., Kahn, R. A. & Ringe, D. Structure of the human ADP-ribosylation factor 1 complexed with GDP. *Nature* **372**, 704–708 (1994).
25. Madej, T., Gibrat, J. F. & Bryant, S. H. Threading a database of protein cores. *Proteins Struct. Funct. Genet.* **23**, 356–369 (1995).
26. Schalk, I. *et al.* Structure and mutational analysis of Rab GDP-dissociation inhibitor. *Nature* **381**, 42–48 (1996).
27. Keep, N. H. *et al.* A modulator of rho family G proteins, rhoGDI, binds these G proteins via an immunoglobulin-like domain and a flexible N-terminal arm. *Structure* **5**, 623–633 (1997).
28. Kolanus, W. *et al.* α L β 2 integrin/LFA-1 binding to ICAM-1 induced by cytohesin-1, a cytoplasmic regulatory molecule. *Cell* **86**, 233–242 (1996).
29. Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
30. De la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).

Acknowledgements. We thank the staff at the LURE for making outstation W32 available to us; M. Roth at the ESRF for beamtime assistance at outstation D2AM; S. Bryant and T. Madej for the Protein Data Bank survey with VAST; J.-F. Gibrat for sequence-threading calculations; S. Paris for unpublished observations; and M. Chabre and J. Janin for discussions. This work was supported by the Association pour la Recherche contre le Cancer and by Zéneca Pharma (France). M.M. was supported by an EMBO fellowship.

Correspondence and requests for materials should be addressed to J.C. (e-mail: cherfils@lebs.cnrs-gif.fr). Coordinates have been deposited in the Brookhaven Data Bank under accession code 1PBV.

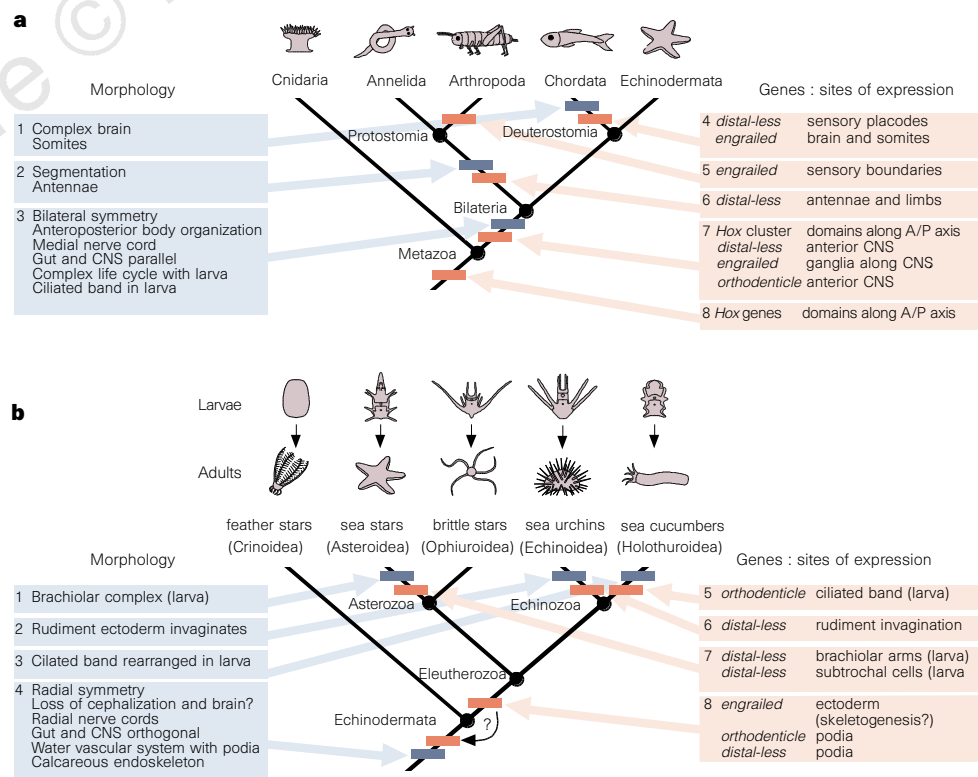
erratum

Radical alterations in the roles of homeobox genes during echinoderm evolution

Christopher J. Lowe & Gregory A. Wray

Nature **389**, 718–721 (1997)

The labelling of Fig. 1 of this Letter was incomplete in several places. The corrected figure is reproduced below.



Antibodies and immunology

This group of products includes a collagen-based kit for identifying and quantifying megakaryocytic progenitors, flow cytometry reagents, an antigen retrieval kit and a permeabilization reagent to help detect intracellular antigens.

Antigen retrieval kit

From BioGenex

This antigen retrieval technique restores immunoreactivity of antigens in formalin-fixed tissue sections

The technique consists of heating the tissue sections in a microwave oven in the presence of retrieval solutions. This accessory kit consists of slide holders and slide baths, and to accommodate microwave heating, the baths and holders are made of heat-stable thermoplastic resins, which may be used with or without a pressure cooker. Three antigen retrieval solutions covering a wide pH range are available for use with the accessory kit, namely Citra, AR-IO and Glyca.

Reader Enquiry No. 100

Architect i2000

From Abbott Laboratories

Flexible instrumentation that helps to meet immunoassay and chemistry testing needs

As a monomeric unit, the i2000 is said to be capable of performing up to 200 tests per hour and processing up to 1,600 tests per eight-hour shift. It provides continuous, random sample access and accommodates 125- or 250-primary-tube loadup for maximum walk-away capability. The i2000 can be upgraded by combining up to four i2000 modules through one operator, thus creating an i8000. The i8000 is stated to perform up to 800 tests per hour with an onboard reagent capacity of 100 different refrigerated immunoassays and 50,000 tests. The i2000 launch provides 26 assays and features Chemiflex technology, which incorporates chemiluminescent technology and flexible protocols. The Architect family comprises three product lines, including the i series (immunoassay), the c series (chemistry) and the ci series (chemistry and immunoassay).

Reader Enquiry No. 101

Immunology

Immulate IL-8 assay

From DPC

A fully automated, random-access interleukin-8 (IL-8) assay

Designed for research into inflammatory reactions, regulatory functions, pathological conditions and disease states, this assay has a working range of 20–10,000 pg ml⁻¹, with a sensitivity of 6.2 pg ml⁻¹, and run-times of about 40 minutes. The assay has been calibrated against NIBSC 89/520 and is specific for IL-8, showing no cross reactivity with other chemokines. The system is



DPC's Immulate assays help study disease states.

capable of handling up to 120 tests per hour and automatically controls all aspects of system management. There are more than 75 immunoassays available in this range.

Reader Enquiry No. 102

rProtein L

From Actigen

Three conjugated forms of this recently introduced affinity binding protein

rProtein L is now supplied conjugated with horseradish peroxidase (HRP), alkaline phosphatase (AP) or biotin, as well as being available in the unconjugated form. All are available in 1-mg quantities in solution. Applications for rProtein L include the purification and detection of immunoglobulins containing κ light chains, purification of Fab and scFv fragments and the purification of monoclonal antibodies directly from media supplemented with BSA or FCS. Coupling rProtein L to HRP and AP enables easy detection in western blotting and ELISA systems, for example, whereas biotin-conjugated rProtein L has applications in the magnetic separation and detection of antibodies. This is a recombinant protein derived from the bacterium *Peptostreptococcus magnus*. It binds immunoglobulins primarily through κ light chain interactions without interfering with antigen binding.

Reader Enquiry No. 103

Mega Cult-C

From Stem Cell

A new collagen-based kit for identifying and quantifying megakaryocytic progenitors

This procedure is stated to take less than two weeks from start to finish and is based on culture fixation and staining performed directly on a standard-size slide. A serum-free growth medium is used to assay for human mega-

karyocytic progenitor cells, as fetal bovine serum contains inhibitors to growth. The identification, cloning and availability of recombinant cytokines, notably thrombopoietin, has permitted the optimization of conditions for the proliferation and differentiation of CFU-Mk (colony forming unit-megakaryocyte). Furthermore, the use of a compatible gelling agent allowing clonal progeny of a single cell to stay together is essential for colony formation. In this system, collagen acts as a stable gel matrix to support growth and development, and allows *in situ* immunocytochemical staining of colonies within the dehydrated gel. The staining system identifies three colony types: pure Mk, mixed Mk/non-Mk and non-Mk. A permanent record of assay results can be stored in standard slide boxes.

Reader Enquiry No. 104

ActiCyte-TC

From BioErgonomics

A complete turnkey system for isolation and activation of human peripheral blood T cells

Each ActiCyte-TC kit includes PrepaCyte-SC medium for rapid isolation of T cells, VitaLyse for removal of residual contaminating erythrocytes, a medium for T-cell activation via cell-surface receptors and a T-cell basal medium for non-activated cultures. Part of a complete T-cell activation system, the kits should be useful for studying receptor expression and cytokine production and secretion.

Reader Enquiry No. 105

$\beta 2m/A^b$ double knockout mice

From Taconic

Mice that lack expression of both classes of MHC antigens

This animal model may be useful for examining compensation mechanisms in severely immunocompromised animals. These mice combine the characteristics of two individual 'knockouts': the $\beta 2m$ and the A^b (relevant to transplantation, gene therapy and infectious disease studies). Moreover, the $\beta 2m/A^b$ double knockout mouse has low levels of MHC (major histocompatibility complex) class I heavy-chain expression, no detectable MHC class II expression and reduced levels of both CD4⁺ and CD8⁺ T cells. These mice were developed by homologous recombination, are available on an inbred C57BL/6 background and carry an H2b haplotype. They were bred by mating the MHC class I-deficient $\beta 2m$ knockout model with the MHC class II-deficient A^b



A double knockout — Taconic's compromised mouse lacks both MHC antigens.

knockout mouse, which are depleted of CD8⁺ and CD4⁺ T cells, respectively. The B-cell compartment of these animals appears intact, and MHC-deficient mice can mount specific antibody responses when challenged with a T cell-independent antigen.

Reader Enquiry No. 106

Spin cycle

RC12BP centrifuge

From Sorvall

A blood-processing centrifuge system for increased productivity and GMP compliance

Capable of accommodating twelve 500-ml blood bags at a time, this centrifuge is designed for low-speed centrifugation and improved GMP compliance in blood component manufacture. With a top speed of 4,700 r.p.m. (7,333g), this machine uses an accumulated centrifugal effect to ensure reproducible, consistent separations regardless of rotor load differences, and a slow start/stop system, which controls rates of acceleration and deceleration to reduce sample resuspension. There are ten different acceleration and deceleration profiles, that can be customized for minimal sample resuspension, cleaner interfaces and higher product yields. The software improves GMP compliance by allowing up to 16 centrifuges from this series to be networked to a single PC for centralized data recording and process control. Also aiding efficiency is a barcode scanner that matches process data to identification numbers for component traceability.

Reader Enquiry No. 107

Assays and cytometry

Flow cytometry reagents

From Stratech Scientific

A collection of single- and multi-colour products, controls and intracellular cytokines

These products are designed and titred to work with all the major flow cytometry instruments. All five-, four-, three- and two-colour products have been optimized for use with single- or dual-laser configurations. According to the manufacturer, users accustomed to looking at one or two populations could be analysing 16 or more at a time in a single tube. Kits for staining intracellular antigens are also offered. Each kit contains pre-mixed solutions for fixation, permeabilization and washing, plus an easy-to-use and optimized protocol.

Reader Enquiry No. 108

AC133-cell antigen

From Miltenyi Biotec

A monoclonal antibody to a novel glycoprotein antigen

Selectively expressed on the surface of immature haematopoietic progenitor cells, this antigen is not detectable on other blood cells, endothelial cells, fibroblasts or myeloid leukaemia cells. AC133-selected cells include the haematopoietic long-term re-population cells demonstrated in a fetal sheep transplantation model. This mAb provides an alternative to CD34 for the selection and characterization of cells necessary for transplant situations, for studies of *ex vivo* expansion strategies and gene therapy studies. AC133 has been conjugated to the company's MACS MicroBeads for magnetic cell sorting, as well as phycoerythrin for flow cytometric analysis.

Reader Enquiry No. 109

Inno-LIA ANA

From Microgen Bioproducts

A multi-parametric immunoassay to detect and differentiate antinuclear antibodies in serum for autoimmune disease diagnosis

Using a new approach to antinuclear antibody (ANA) testing, this assay system utilizes mainly recombinant antigens, allowing it to provide simultaneously information on antibodies to 10 autoantigens and 14 determinants within three hours. Trials indicated that there is increased sensitivity for the detection of anti-Ro52 and anti-La, more so than with a combination of conventional techniques, and it maintains high disease specificity, says Microgen. New diagnostic insights gleaned from this assay include the previously undetected high frequency of anti-Ro52 antibodies found in polymyositis. Efficient high-throughput testing can be achieved using the optional Auto-LIA automated instrument, and objective reading, quantitation and data management is provided with optional LIA-Scan software.

Reader Enquiry No. 110

BDNF_E_{max}

From Promega

The brain-derived neurotrophic factor (BDNF) immunoassay system now comes in an optimized antibody sandwich format

After an overnight coating of a 96-well plate with anti-BDNF mAb, the specific protein is captured and detected using a BDNF polyclonal antibody (pAb) and a species-specific, anti-IgY antibody conjugated to horseradish peroxidase (HRP). Incubation with a chromogenic substrate causes a colour change, easily measured by a plate reader, that is proportional to the amount of BDNF present. The system can be used to quantify BDNF in tissue culture supernatants, plasma, serum, urine or tissue extracts. It is recommended

for use in human, mouse, rat, rhesus monkey, goat, sheep, pig and horse studies. It typically has less than 3% cross-reactivity with other related neurotrophic factors (NGF, NT-3 and NT-4) at 100 ng ml⁻¹. The assay has a stated minimum sensitivity of 15.6 pg ml⁻¹ of BDNF.

Reader Enquiry No. 111

IntraPrep

From Coulter

A new permeabilization reagent for intracellular antigen detection

Formulated for the exploration of intracellular antigen expression by flow cytometry, this reagent allows access to both cytoplasmic and nuclear antigens while preserving light-scatter characteristics and surface-antigen expression. The complete kit offers ready-to-use reagents, formulated to reduce nonspecific staining, and a fast, simple procedure for sample preparation.

Reader Enquiry No. 112

On the web

Cambridge Research Biochemicals offers its antibody guide. The guide can be found at <www.ZenecaLSM.co.uk/crb>.

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

serotec

...better products, better service, better quality...

• Antibodies to Human CDs • Antibodies to Mouse CDs • Antibodies to Rat CDs • Antibodies to Ovine CDs • Antibodies to Bovine CDs • Veterinary Research Antibodies • Antibodies to Hormones • Antibodies to Microbial Antigens • Assays and Reagents • Adjuvant Molecule Products	• Cytokine & Growth Factor Reagents • Cytoskeletal Component Reagents • Flow Cytometry Reagents • Histochemistry Reagents • Cell Signalling Reagents • Neurological Reagents • Reproductive Biology Reagents • Tissue Culture Reagents
---	---

TECH. RESEARCH LAB. CHLT

Services: 22 Boulevard, Marais Lippard, 92100, France (CDs) 01 46 18 18 18
Tel: +33 (0)1 46 18 18 18 Fax: +33 (0)1 46 18 18 18
Technical Enquiries Tel: +33 (0)1 46 18 18 18
Order Hotline Tel: +33 (0)1 46 18 18 18
www.serotec.co.uk

READER ENQUIRY NO. 23

Custom antibody production

Biotechnology Planning - production - after sales support

chicken egg yolk antibodies
- fast - cost effective - very high yield -

rabbit antibodies

monoclonal antibodies
- development and production -

NEW - recombinant antigen service -

Ask for our catalog: Davids Biotechnologie
Karl Anselm Str. 1A, D-93051 Regensburg
Germany, email: Davids.Bio@T-online.de
Phone: +49(0)171/4900308 Fax/Phone: +49(0)941/948228

READER ENQUIRY NO. 26